★ Sampling Theory:

→ practical: 01 09/07/2025

→ sample selection methods: ① Cumulative method

② Lahiri's method.

★ practical - 01 (PPSWR) 16/07/2025

→ formulae:

① unbiased Estimator of population total Y.

$\hat{Y}_{pps} = \frac{1}{n} \sum_{i=1}^{n} y_i/p_i$, where $p_i = \frac{x_i}{X}$, $i = 1, 2, \dots, N$

② Variance of Estimator $\hat{Y}_{pps}$

$$v(\hat{Y}_{pps}) = \frac{1}{n(n-1)} \left(\sum_{i=1}^{n} \frac{y_i^2}{p_i^2} - n \hat{Y}_{pps}^2 \right)$$

③ Standard Error of $\hat{Y}_{pps}$

$$se(\hat{Y}_{pps}) = \sqrt{v(\hat{Y}_{pps})}$$

④ Confidence interval for populational total.

$$C.I. = \hat{Y}_{pps} \pm 2 \cdot se(\hat{Y}_{pps})$$

⑤ Relative Efficiency of PPSWR with SRSWR

$$RE = \frac{V(\hat{Y})}{V(\hat{Y}_{pps})} \times 100 \quad , \quad \text{where } V(\hat{Y}) = \frac{N^2 \cdot \sigma^2}{n}$$

$$\sigma^2 = \frac{1}{N} \left(\sum_{i=1}^{N} Y_i^2 - N \bar{Y}^2 \right)$$

another formula for $V(\hat{Y}) = \frac{1}{n} \left(N \sum_{i=1}^{N} Y_i^2 - Y^2 \right)$

Sampling Theory

Practical-02 (PPSWOR) 23/07/2025

① Desraj's ordered Estimator is (for sample size 2)

$$\hat{Y}_{DR} = \frac{1}{2}\left[\frac{y_1}{P_1}(1+P_1) + \frac{y_2}{P_2}(1-P_1)\right]$$

② Variance of Des-Raj's order Estimator.

$$V(\hat{Y}_{DR}) = \frac{(1-P_1)^2}{4}\left[\frac{y_1}{P_1} - \frac{y_2}{P_2}\right]^2$$

③ Standard Error for Des-Raj's Estimator.

$$SE(\hat{Y}_{DR}) = \sqrt{V(\hat{Y}_{DR})}$$

$$P_i = x_i/x$$

→ For selecting unit in the sample, we choose pair of random numbers (i, R) such that $1 \le i \le 14$ & $1 \le R \le M = 988$.

Thus village : 02 is selected in the sample.

Since the sampling procedure is without Replacement The selected unit (number 2 is not replaced back) Number of units left in the population, are 13.

Now for selecting next unit in the sample we choose another pair of random number such that $i \le 1 \le 13$ & $1 \le R \le M = 988$. we select (8, 17)

i.e. $17 \le 309 = x_8$, 8^{th} village is selected in the sample.

Thus sample selected consist of villages 2nd & 8^{th}

→ During the survey these selected villages will be found to have number of pet Animals as 690 & ~~...~~ 680

③ Murthy's Estimator is $\hat{Y}_M$

$$\hat{Y}_M = \frac{1}{2-P_1-P_2}\left[\frac{y_1}{P_1}(1-P_2) + \frac{y_2}{P_2}(1-P_1)\right]$$

④ Variance using Murthy's Estimator

$$V(\hat{Y}_M) = \frac{(1-P_1)(1-P_2)(1-P_1-P_2)}{(2-P_1-P_2)^2}\left[\frac{y_1}{P_1} - \frac{y_2}{P_2}\right]^2$$

for Q.1) (b) $\hat{Y}_M = 9392.9280$

$V(\hat{Y}_M) = 1030911.617$

⑤ Horvitz Thompson Estimator

$$\hat{Y}_{HT} = \sum_{i=1}^{n} \frac{Y_i}{\pi_i}, \quad \text{where, } \pi_i = P_i\left[\sum_{j=1}^{N}\frac{P_j}{1-P_j} + 1 - \frac{P_i}{1-P_i}\right]$$

π_i = inclusion probability
j = Total popln
i = sample.

⑥ Variance for Horvitz Thompson

$$V(\hat{Y}_{HT}) = \sum_{i\in S}\left(\frac{1}{\pi_i}-1\right)\frac{y_i^2}{\pi_i} + \sum\sum_{i\neq j\in S}\left[\frac{\pi_{ij}}{\pi_i\pi_j}-1\right]\frac{y_i y_j}{\pi_{ij}}$$

, where $\pi_{ij} = P_i P_j\left[1/1-P_i + \frac{1}{1-P_j}\right]$

* Alternate Expression for variance of HT.

$$V_{YG}(\hat{Y}_{HT}) = \sum_{i=1}^{n} \sum_{j>i}^{n} \frac{(\pi_i \pi_j - \pi_{ij})}{\pi_{ij}} \left[\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right]^2$$

Q.3) $V_{YG}(\hat{Y}_{HT}) = 1025269.035$

Multi Variate Analysis - 1 (27/07/2025)

Practical - 01 : Estimation of Conditional distn, Joint distn, partial & multiple correlation.

★ Test procedure:

→ let us Consider two sets of random variables $x_1, x_2, \dots, x_q$ & $x_{q+1}, x_{q+2}, \dots, x_p$ forms the vectors

$$\underline{x}_{(1)} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_q \end{pmatrix}, \quad \underline{x}_{(2)} = \begin{bmatrix} x_{q+1} \\ x_{q+2} \\ \vdots \\ x_p \end{bmatrix}$$

These variables forms the random vector $\underline{x}$ and

$$\underline{x} = \begin{pmatrix} \underline{x}_{(1)} \\ \underline{x}_{(2)} \end{pmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}$$

further let us assume that x is distributed accordingly to $X \sim N_p(\underline{\mu}, \Sigma)$, $\Sigma > 0$, where,

$$\underline{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \begin{bmatrix} \underline{\mu}_{(1)} \\ \underline{\mu}_{(2)} \end{bmatrix}$$

& $$\Sigma = (\sigma_{ij}) = \begin{array}{cc} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} & \begin{array}{c} q \\ p-q \end{array} \\ \begin{array}{cc} q & p-q \end{array} & \end{array}$$

Multi Variate Analysis - 1

Practical - 01 : Estimation of Conditional distⁿ, Joint distⁿ, partial & multiple Correlation.

★ **Test procedure!**

→ let us Consider two sets of random variables $x_1, x_2, \ldots, x_q$ & $x_{q+1}, x_{q+2}, \ldots, x_p$ forms the vectors

$$\underline{x}_{(1)} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_q \end{pmatrix}, \quad \underline{x}_{(2)} = \begin{bmatrix} x_{q+1} \\ x_{q+2} \\ \vdots \\ x_p \end{bmatrix}$$

These variables forms the random vector $\underline{x}$ and

$$\underline{x} = \begin{pmatrix} \underline{x}_{(1)} \\ \underline{x}_{(2)} \end{pmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}$$

further let us assume that x is distributed accordingly to $x \sim N_p(\underline{\mu}, \Sigma)$, $\Sigma > 0$, where,

$$\underline{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \begin{bmatrix} \underline{\mu}_{(1)} \\ \underline{\mu}_{(2)} \end{bmatrix}$$

& $$\Sigma = (\sigma_{ij}) = \begin{array}{cc} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} & \begin{array}{c} q \\ p-q \end{array} \\ \begin{array}{cc} q & \quad p-q \end{array} & \end{array}$$

★ **Conditional distribution:** The Conditional dist^n of X_1 given X_2 $\underline{x}_{(1)}$ given $\underline{x}_{(2)} = \underline{x}_{(2)}$ is q variable normal with

mean $\underline{\mu}_{1.2} = \underline{\mu}_{(1)} + \Sigma_{12} \Sigma_{22}^{-1} (\underline{x}_{(2)} - \underline{\mu}_{(2)})$

→ Covariance $= \Sigma_{11.2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$

→ Note that Covariances does not depend on the value $\underline{x}_{(2)}$ of Conditional Variables.

The matrix $\beta = \Sigma_{12} \Sigma_{22}^{-1}$ is matrix of regression coefficients of $\underline{x}_{(1)}$ on $\underline{x}_{(2)}$

→ The $\underline{\mu}_{1.2}$ is often called regression function.

→ The matrix $\Sigma_{11.2}$ is called the residual Var-Cov matrix.

→ The Correlation Coefficient bet^n i^{th} & j^{th} Variables defined as

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}$$

→ The matrix, $P = (\rho_{ij})$ with $\rho_{ii} = 1$ is called Correlation matrix

★ **Partial Correlation:** $\rho_{ij.q+1,\ldots,p} = \dfrac{\sigma_{ij.q+1,\ldots,p}}{\sqrt{\sigma_{ii.q+1,\ldots,p}\,\sigma_{jj.q+1,\ldots,p}}}$

is the partial Correlation bet^n x_i & x_j holding $x_{q+1},\ldots,x_p$ variables fixed.

→ Note that simple Correlation is defined from the elements of the Variance Covariance matrix σ whereas the partial correlation is defined from the element of the residual Variance-Covariance matrix.

★ Multiple Correlation Coefficient :

→ The maximum Correlation betn X_i & the linear Combination $\underline{\alpha}' \underline{X}_{(2)}$ is called the multiple Correlation Coefficient betn $\underline{X}_i$ & $\underline{X}_{(2)}$

$$\rho_{i(q+1,\ldots,p)} = \sqrt{\frac{\underline{\sigma}'_{(i)} \Sigma_{22}^{-1} \underline{\sigma}_{(i)}}{\sigma_{ii}}}$$

$$= \sqrt{\frac{\Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}}{\Sigma_{11}}}$$

★ Joint distribution:

$$P_r\{(X,Y) \in E\} = \iint_E f(x,y)\,dx\,dy.$$

Q.2) $X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}_{3\times1} \sim N_3(\mu, \Sigma)$

$\mu = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}_{3\times1}$, $\Sigma = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 3 & 0 \\ 1 & 0 & 1 \end{bmatrix}_{3\times3}$

$Y_1 = X_1 + X_2 + X_3$

$Y_2 = X_1 - X_2$ $\Rightarrow Y = AX$, where $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \end{bmatrix}_{2\times3}$

Mean of Y is, $\mu_Y = E(Y) = A\mu = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \end{bmatrix}_{2\times3} \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}_{3\times1}$

$$\therefore E(Y) = \begin{bmatrix} 4 \\ 1 \end{bmatrix}_{2\times1}$$

27/07/25

[illegible]

Faculty of [illegible]

M[illegible] Analysis -I (M.Sc. Sem 1)

Practical 1- Estimation of Conditional Distribution, Joint Distribution, Partial [illegible]
[illegible] Correlation

Q-1) Following is a summary of [illegible] observations of a random vector $X = (x_1, [illegible])$

$\bar{X} =$ (151.[illegible], 1[illegible].24, [illegible], 1[illegible].54)

$$S = \begin{pmatrix} 54.36060 & [illegible] & [illegible] & 51.31[illegible] \\ & [illegible] & [illegible]17 & 56.5400 \\ & & [illegible]33 & 69.[illegible] \\ & & & 100.8067 \end{pmatrix}$$

i) Obtain the conditional di[illegible] of (x_1, x_3) given x_2 = 150

ii) Compute the simple correlation coefficient between x_1 and [illegible]

iii) Compute the partial correlation coefficient between x_1 and [illegible] where [illegible] control.

iv) Compute the maximum correlation coefficient that can be obtained between [illegible] and linear combination of x_1, [illegible]

Q-2) Let $X \sim N_3(\mu, \Sigma)$ where $\mu' = (2, 1, [illegible])$ and Σ = [illegible] 1 / [illegible] 0 / [illegible] 1

Find the joint distribution of $Y_1 = x_1 + x_2 + x_3$ and $Y_2 = [illegible] - x_2$

Q-3) Let $X \sim N_3(\mu, \Sigma)$ where $\mu' = (3, 2, 4)$ and Σ = [illegible] 2 1 / [illegible] 5 3 / 1 [illegible] 6

Find the joint distribution of $Y_1 = 2x_1$ [illegible] $Y_2 = [illegible] x_3$

****

→ Covariance matrix of Y is,

$$\Sigma_Y = A \Sigma A' = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \end{bmatrix}_{2\times3} \begin{bmatrix} 2 & 1 & 1 \\ 1 & 3 & 0 \\ 1 & 0 & 1 \end{bmatrix}_{3\times3} \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 0 \end{bmatrix}_{3\times2}$$

$$= \begin{bmatrix} 4 & 4 & 2 \\ 1 & -2 & 1 \end{bmatrix}_{2\times3} \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 0 \end{bmatrix}_{3\times2}$$

$$= \begin{bmatrix} 10 & 0 \\ 0 & 3 \end{bmatrix}_{2\times2}$$

→ Covariance matrix of Y is,

$$\Sigma_y = A \Sigma A' = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \end{bmatrix}_{2\times 3} \begin{bmatrix} 2 & 1 & 1 \\ 1 & 3 & 0 \\ 1 & 0 & 1 \end{bmatrix}_{3\times 3} \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 0 \end{bmatrix}_{3\times 2}$$

$$= \begin{bmatrix} 4 & 4 & 2 \\ 1 & -2 & 1 \end{bmatrix}_{2\times 3} \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 0 \end{bmatrix}_{3\times 2}$$

$$= \begin{bmatrix} 10 & 0 \\ 0 & 3 \end{bmatrix}_{2\times 2}$$

* practical - 02 MLE for variance Covariance matrix :

27/07/2025

$\rightarrow$ MLE of $\hat{\mu} = \bar{x} = \frac{1}{N} \sum_{j=1}^{N} x_j = N^{-1} X E_{N1}$

$\rightarrow$ MLE of Σ is $\hat{\Sigma} = N^{-1} \sum_{j=1}^{N} (\underline{x}_j - \bar{x})(\underline{x}_j - \bar{x})'$

$$= N^{-1} X [I_N - N^{-1} E_{NN}] X'$$

$$E_{N1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{N \times 1}$$

$$E_{NN} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}_{N \times N}$$

Q.1)

Mean = $y_1 \mu_1 + y_2 \mu_2$

Variance = $V(x_1 - x_2)$

Q.2)

mean = take averages.

Var-Cov. matrix = varP Covariance.p

$\rightarrow$ rupali.k-stat@msubaroda.ac.in

Practical - 05 Sampling theory (Random Group Method)

23/08/2025

$Y_{ij} \rightarrow$ Total yield, for the j^{th} village in the i^{th} random group. in 1000 quintals

$X_{ij} \rightarrow$ Area under wheat crops (in hectare) for the j^{th} village in the i^{th} random Group.

$P_{ij} \rightarrow$ Initial probability of selection for the j^{th} village in the i^{th} random group.

Sol, Q.1) Since we have to select a sample of size 4, four random Groups need to be Constructed. Thus first two groups are formed of 4 units each and the remaining two groups consist of 5 units each.

We choose 18 random numbers from 1 to 18 without replacement which amount to arranging the given population in a random order.

$$\hat{Y}_{RHC} = \frac{y_1}{p_1}\phi_1 + \frac{y_2}{p_2}\phi_2 + \frac{y_3}{p_3}\phi_3 + \frac{y_4}{p_4}\phi_4$$

$$v(\hat{Y}_{RHC}) = \left[\frac{N_1^2 + N_2^2 + N_3^2 + N_4^2 - N}{N^2 - (N_1^2 + N_2^2 + N_3^2 + N_4^2)}\right]\left[\left(\frac{y_1}{p_1} - \hat{Y}_{RHC}\right)^2\phi_1 + \left(\frac{y_2}{p_2} - \hat{Y}_{RHC}\right)^2\phi_2 + \left(\frac{y_3}{p_3} - \hat{Y}_{RHC}\right)^2\phi_3 + \left(\frac{y_4}{p_4} - \hat{Y}_{RHC}\right)^2\phi_4\right]$$

$$C.I. = \hat{Y}_{RHC} \pm 2\, S.E\,(\hat{Y}_{RHC})$$

Practical-04 Stratified PPSWR

$$\hat{y} = \sum_{i=1}^{k} \hat{y}_i \quad , \quad \text{where } k \text{ is no. of words.}$$

and

$$\hat{y}_i = \sum_{j=1}^{N_i} \frac{x_{ij}}{m_i} \cdot \sum_{j=1}^{m_i} y_{ij}/x_{ij} \quad , \quad \begin{array}{l} i = 1, 2, 3 \\ \& \; j = 1, 2, \ldots, 10 \end{array}$$

$$v(\hat{y}) = v\left(\sum_{i=1}^{k} \hat{y}_i\right) = \sum_{i=1}^{k} v(\hat{y}_i)$$

where , $v(\hat{y}_i) = \frac{1}{m_i(m_i-1)}\left[\left(\sum_{i=1}^{N_i} x_{ij}\right)^2 \sum_{j=1}^{m_i} \frac{y_{ij}^2}{x_{ij}^2} - m_i \hat{y}_i^2\right]$

* Practical : 03 Wishart Distribution

→ let A be a wishart matrix

$A = XX' = BYY'B' = BB'$

→ standard pdf of A becomes,

$$w(n, I) = K(p, n)\, |A|^{\frac{n-p-1}{2}} \exp\left\{-\frac{1}{2} \text{Tr.} A\right\}, \quad A > 0$$

$$\text{where} \quad K(p, n) = \frac{1}{2^{np/2}\, \pi^{\frac{p(p-1)}{4}} \prod_{i=1}^{p} \sqrt{\frac{n-1+i}{2}}}$$

$K(p, n)$ is normalized Constant.

$\{ Y$ is orthogonal, $YY' = Y'Y = I$ or $Y' = Y^{-1} \}$

* Properties of wishart Distribution:

1. If $A \sim W_p(\Sigma, n)$ & $B : q \times p$ then $BAB' \sim W_p(B\Sigma B', n)$

we know that if $X \sim N_p(0, \Sigma)$ then $A = XX' \sim W_p(\Sigma, n)$

Here, $BAB' = BXX'B' = YY'$ if $Y = BX$.

But $Y \sim N_p(0, B\Sigma B')$

$\therefore YY' \sim W_p(B\Sigma B', n)$ where n is d.f.

2. If $A \sim W_p(\Sigma, n)$ then $|A|$ is distributed as the product of p independent χ^2 variates $|\Sigma|$ with d.f. $n, n-1, \ldots, n-(p-1)$ respectively.

3. If $A \sim W_p(\Sigma, m)$ and $B \sim W_p(\Sigma, n)$ are independent and if $m \geq p$, $n \geq p$ then $\phi = |A^{-1}B| = \frac{|B|}{|A|}$ is proportional to the product of p independent F variables where ith product has $(n-i+1)$ & $(m-i+1)$ d.f.

$\phi = |A^{-1}B| = \frac{|B|}{|A|}$ is proportional to the product of p independent F variables where the i^{th} product has $(n-i+1)$ & $(m-i+1)$ d.f. i.e. $\phi = |A^{-1}B| = \frac{|B|}{|A|} \propto \prod_{i=1} F_\alpha$

-> $A \sim W_p(I, m)$ and $B \sim W_p(I, n)$ are independent, $m \geq p$ then, $\lambda = \frac{|A|}{|A+B|} = |I + A^{-1}B|^{-1}$

has a will's λ distribution with parameters (p, m, I) & notation $\lambda(p, m, n)$

Q.1}

$$S = \frac{1}{N-1} \Sigma (X - \bar{X})(X - \bar{X})'$$

$$S(N-1) = XX'$$

Compute BAB'

Verify checking mean, $\mu = E(BAB') = n(BSB')$
where $S \approx \Sigma$

$\because$ If $A \sim W_p(\Sigma, n)$ then $E(A) = n\Sigma$ &

$$E(A^{-1}) = \frac{1}{n-p-1} \Sigma^{-1}$$

Provided $n-p-1 > 0$

Q.2> $A \sim W_p(\Sigma, n)$. $\phi = \frac{|A|}{|\Sigma|} \sim \prod_{i=1}^{p} \chi_{(\alpha, d.f.)}$

chisq.INV.RT generate 4 random no.
Then using it apply chisq.INV.
then.

Q.3> $p = 3$, $N_1 = 151$, $n_1 = 150$, $N_2 = 101$, $n_2 = 100$

$A = (N_1 - 1) S_A$, $B = (N_2 - 1) S_B$

$A \sim W_3(\Sigma, 150)$, $B \sim W_p(\Sigma, 100)$

$$\frac{|B|}{|A|} \propto \prod_{i=1}^{p} F_{(n-1+i, m-i+1)}$$

$$\frac{|B|}{|A|} = C \prod_{i=1}^{p} F, \quad C = \prod_{i=1}^{p} \frac{(n-i+1)^{-1}}{(m-i+1)^{-1}}$$

16/07/2025

(A)

Maharaja Sayajirao Uni[illegible] of [illegible]

Faculty of [illegible]

Department of Statisti[illegible]

Practical No. [illegible]

[illegible]

Q1. In a primary [illegible] the aim is to estimate the average [illegible] of [illegible] with the help of a sample of students. there are in all [illegible] respective strength of students as 47,30,40,60,45,30,[illegible] to select a [illegible] sections for selection students [illegible] considering the student of the section as a size of measure, [illegible] proportional to [illegible] with replacement sample.

Q2. Use the data [illegible] example 1 for illustrate the procedure of [illegible] sample [illegible] method with probability proportional to [illegible] replacement sampling scheme.

Q3. The [illegible] area of 69 villages of [illegible] given in [illegible] a sample of 10 villages, using PPS [illegible] sampling, [illegible] area as the size measure. Therefore, [illegible] number of [illegible] block along with its standard error, and [illegible] limits on the [illegible]. Also compute relative efficiency of PPS [illegible]

Villa[illegible]	[illegible]ube wells	Villag[illegible]	[illegible]
3	70	6	
32	97	23	
35	116	63	
[illegible]	5[illegible]1	9	
36	113	53	