Vivekanand College Kolhapur (Empowered Autonomous)

B.Sc. II Semester III

Surprise Test

Paper: Statistical Methods I

Date: 29/01/2025 **Total Marks: 20M**

Instruction:

Attempt all questions

Q 1. Define residual & order of residual. State & prove three properties of residual. **10M**

Q 2. State the equation of regression planes. State the formulas of all partial regression equations **5*2=10**

Q3. With usual notations show that, Cov $(X_1, e_{1.23}) = \sigma_1{}^2 - \sigma_{1.23}{}^2$ where $e_{1.23}$ is estimated value.

Name = Poonam Sachin Godane
RollNo = 7739
subject = statistics

29/11/2

PAGE NO.
DATE

20/20

Q.1] Define Residual and Order of Residual. state and Prove Three Properties of Residual.

Ans. • Residual = The Difference between The Acutual value and it's estimated value is called as Residual.

• order of Residual = The order of Residual is The Number of secondary subscripts in it's notation, is called as order of Residual.

• Property 1 :

• statement = The sum of The Product of any Residual of order zero with the Residual of Highest order zero Provided, The subscripts of The former occurs among The secondary subscripts of The Latter.

i.e $\Sigma X_2 X_{1.23} = 0$

• Proof :- Given,

$\Sigma X_2 X_{1.23}$

Now $\Sigma X_2 (X_1 - b_{12.3} X_2 - b_{13.2} X_3)$

$= \Sigma X_1 X_2 - b_{12.3} \Sigma X_2^2 - b_{13.2} \Sigma X_2 X_3$

$= n\sigma_1\sigma_2 r_{12} - b_{12.3}\, n\sigma_2^2 - b_{13.2}\, n\sigma_2\sigma_3 r_{23}$

$= n\sigma_1\sigma_2 r_{12} - \left(\frac{-\sigma_1}{\sigma_2}\frac{R_{12}}{R_{11}}\right) n\sigma_2^2 - \left(\frac{-\sigma_1}{\sigma_3}\frac{R_{13}}{R_{11}}\right) n\sigma_2\sigma_3 r_{23}$

PAGE NO.
DATE :

$$n\sigma_1\sigma_2\left[r_{12}+\frac{R_{12}}{R_{11}}+\frac{R_{13}}{R_{11}}r_{23}\right]$$

$$\therefore \frac{n\sigma_1\sigma_2}{R_{11}}\left[r_{12}R_{11}+R_{12}+R_{13}r_{23}\right]$$

$$=\frac{n\sigma_1\sigma_2}{R_{11}}\left[r_{12}(1-r_{23}^2)+(r_{13}r_{23}-r_{12})+(r_{23}r_{12}-r_{13})r_{23}\right]$$

$$=\frac{n\sigma_1\sigma_2}{R_{11}}\left[\cancel{r_{12}}-\cancel{r_{12}r_{23}^2}+\cancel{r_{13}r_{23}}-\cancel{r_{12}}+\cancel{r_{12}r_{23}^2}-r_{13}\cancel{r_{23}}\right]$$

$$=\frac{n\sigma_1\sigma_2}{R_{11}}\times 0$$

$$=0$$

$\therefore \Sigma X_2 X_{1.23}=0$

Hence Property is Proved.

2) Property 3 :

- Statement = The sum of The Product of any Two Residual is zero, if All the subscripts of The one occurs among The secondary subscripts of The other. i.e $\Sigma X_{1.2} X_{3.12}=0$

- Proof = Given,

$$\Sigma X_{1.2} X_{3.12}$$
$$=\Sigma X_{3.12}(X_1-b_{12}X_2)$$
$$=\Sigma X_1 X_{3.12}-b_{12}\Sigma X_2 X_{3.12}$$
$$=0 \quad \text{— by Property (1)}$$

hence $\Sigma X_1 X_{3.12}=0$

PAGE NO.

DATE :

Simillary, we Prove That

$\cancel{\Sigma X3.1 X1.2}$, $\Sigma X3.1\ X2.13 = 0$

3) Property 2

Statement = $\Sigma X1.2\ X1.23 = \Sigma X1\ X1.23 = \Sigma X1.23^2$

i.e The sum of Product of Any Residual of order zero with Residual of highest order zero Provided, The subscripts of the Residual where we can Add or substract the any Residual.

• Proof : $\Sigma X1.23^2$

$= \Sigma X1.23\ X1.23$

$= \Sigma X1.23\ (X1 - b12.3 X2 - b13.2 X3)$

$= \Sigma X1\ X1.23 - b12.3 \Sigma X2\ X1.23 - b13.2 \Sigma X3\ X1.23$

$= \Sigma X1\ X1.23$ —— by Property (1)

--- (1)

also,

$\therefore \Sigma X1.2\ X1.23$

$= \Sigma X1.23\ (X1 - b12 X2)$

$= \Sigma X1\ X1.23 - b12 \Sigma X2\ X1.23$

$= \Sigma X1\ X1.23$ — by Property (1)

— (2)

10

from eq^n (1) and (2) we Prove That

$\Sigma X1.23^2 = \Sigma X1.2\ X1.23$

$\therefore \Sigma X1.23^2 = \Sigma X1.2\ X1.23 = \Sigma X1\ X1.23$

Hence Property is Proved.

PAGE NO.
DATE :

Q.2] Give The Equations of Regression Planes and formula's of All Partial Regression Coefficients.

Ans.

1) The Equation of Regression Plane of X_1 on X_2 and X_3 is Given by,

$$\therefore \frac{R_{11}}{\sigma_1} X_1 + \frac{R_{12}}{\sigma_2} X_2 + \frac{R_{13}}{\sigma_3} X_3 = 0$$

2) The Equation of Regression Plane of X_2 on X_1 and X_3 is Given by,

$$\therefore \frac{R_{21}}{\sigma_1} X_1 + \frac{R_{22}}{\sigma_2} X_2 + \frac{R_{23}}{\sigma_3} X_3 = 0$$

3) The Equation of Regression Plane of X_3 on X_1 and X_2 is Given by,

$$\therefore \frac{R_{31}}{\sigma_1} X_1 + \frac{R_{32}}{\sigma_2} X_2 + \frac{R_{33}}{\sigma_3} X_3 = 0$$

- Formulae's of Partial Regression Coefficients

1) Partial Regression Coefficients of X_1 on X_2 and X_3 are constants which can be denoted by $b_{12.3}$ and is Given by,

$$b_{12.3} = -\frac{\sigma_1}{\sigma_2} \frac{R_{12}}{R_{11}}$$

2) $$b_{13.2} = -\frac{\sigma_1}{\sigma_3} \frac{R_{13}}{R_{11}}$$

where $b_{13.2}$ are Partial Regression Coefficient of X_1 on X_3 w.r.t. X_2.

3) $b_{21.3} = -\frac{\sigma_2}{\sigma_1} \frac{R_{21}}{R_{22}}$

where $b_{21.3}$ is Partial Regression coefficient of x_2 on x_1 w.r.t. x_3.

4) $b_{23.1} = -\frac{\sigma_2}{\sigma_3} \frac{R_{23}}{R_{22}}$

where $b_{23.1}$ is Partial Regression coefficient of x_2 on x_3 w.r.t. x_1

5) $b_{31.2} = -\frac{\sigma_3}{\sigma_1} \frac{R_{31}}{R_{33}}$

where $b_{31.2}$ is Partial Regression coefficient of x_3 on x_1 w.r.t. x_2.

6) $b_{32.1} = -\frac{\sigma_3}{\sigma_2} \frac{R_{32}}{R_{33}}$

5

Where $b_{32.1}$ is Partial Regression coefficient of x_3 on x_2 w.r.t. x_1.

Q.3] With usuall notations show that, $\mathrm{Cov}(x_1, e_{1.23}) = \sigma_1^2 - \sigma_{1.23}^2$ where $e_{1.23}$ is estimated value of x_1.

Ans.

The equation of Regression Plane x_1 on x_2 and x_3 is Given by,

$$x_1 = b_{12.3} x_2 + b_{13.2} x_3$$

Now $\hat{x_1}$ = The Residual of estimate value of x_1 where known value of x_2.

$\hat{x_1} = e_{1.23}$

Now, The Residual of x_1 on x_2 & x_3 is,

$x_{1.23} = x_1 - \hat{x_1}$

$= x_1 - e_{1.23}$

$\therefore \quad e_{1.23} = x_1 - x_{1.23}$ —— (1)

Now, $cov(x_1, e_{1.23})$

$= E(x_1, e_{1.23}) - E(x_1) \times E(e_{1.23})$

$= E(x_1\, e_{1.23})$ —— ($\because E(x_1) = E(e_{1.23})$

$= 0$ because variables x_1, x_2 & x_3 are measured from their Respective mean)

$\therefore \quad E(x_1\, e_{1.23}) = E(x_1\, e_{1.23})$

$= \frac{1}{n} \sum x_1\, e_{1.23}$

$= \frac{1}{n} \sum x_1 (x_1 - x_{1.23})$ — from eq^n (1)

$= \frac{1}{n} \sum x_1^2 - \frac{1}{n} \sum x_1\, x_{1.23}$

$= \frac{1}{n} \sum x_1^2 - \frac{1}{n} \sum x_{1.23}\, x_{1.23}$ —— by Property (2)

$= \frac{1}{n} \sum x_1^2 - \frac{1}{n} \sum x_{1.23}^2$

$= \sigma_1^2 - \sigma_{1.23}^2$

hence,

$\boxed{cov(x_1, e_{1.23}) = \sigma_1^2 - \sigma_{1.23}^2}$

Hence Proved.

Page No.
Date

* Property 2 :

- statement = The sum of Product of any Residual in which all the secondary subscripts of The first occurs among The secondary subscripts of The second is unaltered. if we delete or any secondary subscript of The first.

i.e $\Sigma x_{1.2} x_{1.23} = \Sigma x_1 x_{1.23} = \Sigma x_{1.23}^2$

- Proof = Given,

$$\Sigma x_{1.23}^2 = \Sigma x_{1.23} x_{1.23}$$

$$= \Sigma x_{1.23} (x_1 - b_{12.3} x_2 - b_{13.2} x_3)$$

$$= \Sigma x_1 x_{1.23} - b_{12.3} \Sigma x_2 x_{1.23} - b_{13.2} \Sigma x_3 x_{1.23}$$

$$= \Sigma x_1 x_{1.23} \quad \text{— by Property (1)}$$

— (1)

also,

$$\Sigma x_{1.2} x_{1.23} = \Sigma x_{1.23} (x_1 - b_{12} x_2)$$

$$= \Sigma x_1 x_{1.23} - b_{12} \Sigma x_2 x_{1.23}$$

$$= \Sigma x_1 x_{1.23} \quad \text{— by Property (1)}$$

— (2)

from eq^n (1) & (2) we get,

$$\Sigma x_{1.23}^2 = \Sigma x_{1.2} x_{1.23}$$

Hence,

$$\Sigma x_{1.23}^2 = \Sigma x_{1.2} x_{1.23} = \Sigma x_1 x_{1.23}$$

Hence Proved.

PAGE NO.
DATE :

Name :- Vinit Nikhil Magdum
Roll No. 7750

$\frac{18}{20}$

Q1 Residual. :-

The difference bet^n actual value and it's estimated value is called as residual.

order of Residual :-

No. of secondary subscript of Residual is called as order of residual.

property 1.

The sum of product of any residual order zero. with any highest order zero provided the subscript occur amoung secondary subscript of latter

$$\sum X_{2} \cdot X_{1\cdot 23} = \sum X_{3} \sum X_{1\cdot 23} = 0$$

proof

$$\sum X_{2} \; X_{1\cdot 23} = \sum X_2 (X_1 - b_{12\cdot 3} X_2 - b_{13\cdot 2} X_3]$$

$$= \sum X_1 X_2 - b_{12\cdot 3} \sum X_2^2 - b_{13\cdot 2} \sum X_2 X_3$$

$$= n \sigma_1 \sigma_2 r_{12} - b_{12\cdot 3}\, n \sigma_2^2 - b_{13\cdot 2}\, n \sigma_2 \sigma_3 r_{23}$$

$$= n \sigma_1 \sigma_2 r_{12} - \left(\frac{-\sigma}{\sigma_2} \frac{R_{12}}{R_{11}} \right) n \sigma_2^2 \cdot \left(\frac{-\sigma}{\sigma_3} \frac{R_{13}}{R_{11}} \right) n \sigma_2 \sigma_3 r_{23}$$

$$= n \sigma_1 \sigma_2 \left[r_{12} + \frac{R_{12}}{R_{11}} + \frac{R_{13}}{R_{11}} + r_{23} \right]$$

PAGE NO.
DATE :

$$= \frac{n\sigma_1\sigma_2}{R_{11}}\left[r_{12} + R_{11} + R_{13} + r_{23}\right]$$

$$= \frac{n\sigma_1\sigma_2}{R_{11}} \times 0$$

$$= 0$$

$\Sigma X_2\, X_{1\cdot 23} = 0$ simillary $X_3\, X_{1\cdot 23} = 0$

property 2 :-

The sum of product of any residual in which all the secondary subscript of first occur amounge secondary subscript of second. if we delet any or all the secondary subscript of first.

$$\Sigma X_{1\cdot 23}\, X_{1\cdot 23} = \Sigma X_{1\cdot 2} X_{1\cdot 23} = \Sigma X^2_{1\cdot 23}$$

proof

$$\Sigma X_{1\cdot 23}\; X_{1\cdot 23} = \Sigma X_{1\cdot 23}\,(X_1 - b_{12\cdot 3} X_2 - b_{13\cdot 2} X_3)$$

$$= \Sigma X_1\, X_{1\cdot 23} - b_{12\cdot 3}\Sigma X_2\, X_{1\cdot 23} - b_{13\cdot 2}\,\Sigma X_3\, X_{1\cdot 23}$$

by property ①

$$\Sigma X_2\; X_{1\cdot 23} = \Sigma X_3\, X_{1\cdot 23} = 0$$

$$X_{1\cdot 23}\;\; X_{1\cdot 23} = \Sigma X_1\;\; X_{1\cdot 23} \quad \text{- - - - -} ①$$

also consider

$$\Sigma X_{12}\;\; X_{1\cdot 23} = \Sigma X_{1\cdot 23}\,(X_1 - b_{12\cdot 3} X_2)$$

$$= \Sigma X_1\cdot X_{1\cdot 23} - b_{12\cdot 3}\Sigma X_2\; X_{1\cdot 23}$$

$$\Sigma X_{1\cdot 2}\;\, X_{1\cdot 23} = X_1\cdot X_{1\cdot 23} \quad \text{. . . . .} ②$$

PAGE NO.
DATE :

from eq^n ① & ②

$$\Sigma x_{1\cdot 23}\, x_{1\cdot 23} = \Sigma x_{1\cdot 2}\, x_{2\cdot 23} = \Sigma x_1^2\, x_{1\cdot 23}$$

property ③

The sum of product of any two residual order zero in which all secondary subscript of first occur amoung secondary subscript of other

$$\Sigma x_{1\cdot 2}\, x_{3\cdot 12} = \Sigma x_{1\cdot 3}\, x_{2\cdot 13}$$

proof

$$\Sigma x_{1\cdot 2}\, x_{3\cdot 12} = \Sigma (x_1 - b_{12\cdot 3}\, x_2)(x_{3\cdot 12})$$

$$= \Sigma x_1\, x_{3\cdot 12} - b_{12\cdot 3}\, \Sigma x_2\, x_{3\cdot 12}$$

by property ① $= 0 - 0$

$= 0$

also we can proved

$$\Sigma x_{1\cdot 3}\, x_{2\cdot 13} = 0$$

10

PAGE NO.

DATE :

Q 2

A] (i) equation of regression plane of x_1 on x_2 and x_3 is given by

$$\frac{R_{11}}{\sigma_1} x_1 + \frac{R_{12}}{\sigma_2} x_2 + \frac{R_{13}}{\sigma_3} x_3$$

(2) equation of regression plane of x_2 on x_1 and x_3 is given by

$$\frac{R_{21}}{\sigma_1} x_1 + \frac{R_{22}}{\sigma_3} x_2 + \frac{R_{23}}{\sigma_3} x_3$$

(3) equation of regression plane of x_3 on x_1 and x_2 is given by

$$\frac{R_{31}}{\sigma_1} x_1 + \frac{R_{32}}{\sigma_2} x_2 + \frac{R_{33}}{\sigma_3} x_3$$

B] Formula for partial regression coefficient.

(1) $b_{1\cdot 23} = \left(\frac{-\sigma}{\sigma_2} \frac{R_{12}}{R_{11}}\right)$

(4) $b_{2\cdot 31} \left(\frac{-\sigma_2}{\sigma_3} \frac{R_{23}}{R_{22}}\right)$

(2) $b_{1\cdot 23} = \left(\frac{-\sigma_1}{\sigma_3} \frac{R_{13}}{R_{11}}\right)$

(5) $b_{32\cdot 1} \left(\frac{-\sigma_3}{\sigma_2} \frac{R_{32}}{R_{33}}\right)$

(3) $b_{21\cdot 3} = \left(\frac{-\sigma_2}{\sigma_1} \frac{R_{21}}{R_{22}}\right)$

(6) $b_{31\cdot 2} \left(-\frac{\sigma_3}{\sigma_2} \frac{R_{31}}{R_{33}}\right)$

PAGE NO.

DATE :

Q3

$cov(x_1, e_{1\cdot 23}) = \sigma_1^2 - \sigma_{1\cdot 23}^2$ where $e_{1\cdot 23}$ is estimated value of x_1

The eq^n of regression x_1 on x_2 & x_3

is

$$x_1 = b_{12\cdot 3}\, x_2 + b_{13\cdot 2}\, x_3$$

Now $\hat{x_1}$ = The residual of estimate value of

$$\hat{x_1} = e_{1\cdot 23}$$

Now

is

The Residual of x_1 on x_2 & x_3

$$x_{1\cdot 23} = x_1 - \hat{x_1}$$
$$= x_1 - e_{1\cdot 23}$$

$$e_{1\cdot 23} = x_1 - x_{1\cdot 23}$$

4

$$cov(x_1 e_{1\cdot 23}) = E(x_1\, e_{1\cdot 23}) - E(x_1)\, E(\cancel{x}_{1\cdot 23})$$

$$= E(x_1\, e_{1\cdot 23})$$

$$= E[x_1\,(x_1 - x_{1\cdot 23})$$

$$= E(x_1^2) - E(x_1\, x_{1\cdot 23})$$

$$= \frac{1}{n}\sum x_1^2 - \frac{1}{n}\sum x_1\, x_{1\cdot 23}$$ by property (2)

$$cov(x_1\, e_{1\cdot 23}) = \sigma_1^2 - \sigma_{1\cdot 23}^2$$

Vivekanand College Kolhapur (Empowered Autonomous)

B.Sc. III Semester V

Surprise Test

Paper: Sampling Theory

Date: 27/08/2024 **Total Marks: 20M**

Instruction:

Attempt all questions

Q 1. In SRSWOR show that $E(s^2) = S^2$ **8M**

Q 2. Attempt any three **(3*4=12)**

1. Define the terms: Population, Sampling unit
2. Explain Sampling Frame
3. Define Sampling. Explain advantages of sampling over census.
4. Explain the term sampling error.

****

Janhavi . S . Mane.

20

Sampling Theory

19/20

8m Q.1) In SRSWOR, show that $E(s^2) = S^2$

OR

Q.1) In SRSWOR, $V(\bar{y}_n) = \frac{N-n}{N} \frac{S^2}{n}$

Q.2) Attempt any 3 . $(4 \times 3 = 12)$

i) Explain sampling frame.

ii) Define sampling. ~~Def~~ Explain advantages of sampling over census method.

iii) Explain the term, sampling errors.

iv) Define population, sampling unit.

→ Q.1) sample mean sq.

$$s^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y}_n)^2}{n-1}$$

$$= \frac{\sum_{i=1}^{n} [y_i^2 - n\bar{y}_n^2]}{n-1}$$

$$= \frac{1}{n-1} \left[\sum_{i=1}^{n} y_i^2 - \cancel{n} \left(\frac{\sum_{i=1}^{n} y_i}{\cancel{n}} \right)^2 \right]$$

$$= \frac{1}{n-1} \left[\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} \right]$$

$$= \frac{1}{n-1} \left[\sum_{i=1}^{n} y_i^2 - \left(\frac{\sum_{i=1}^{n} y_i^2 + \sum_{i \neq j = 1}^{n} y_i y_j}{n} \right) \right]$$

$$= \frac{1}{n-1} \left[\sum_{i=1}^{n} y_i - \frac{\sum_{i=1}^{n} y_i^2}{n} - \frac{\sum_{i \neq j=1}^{n} y_i y_j}{n} \right]$$

$$= \frac{1}{n-1} \left[\frac{n \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} y_i^2}{n} - \frac{\sum_{i \neq j=1}^{n} y_i y_j}{n} \right]$$

$$\therefore\ s^2 = \frac{1}{n-1}\left[\frac{(n-1)\sum_{i=1}^{n} y_i^2}{n} - \frac{\sum_{i\neq j=1}^{n} y_i y_j}{n}\right]$$

$$= \frac{1}{n(n-1)}\left[(n-1)\sum_{i=1}^{n} y_i^2 - \sum_{i\neq j=1}^{n} y_i y_j\right]$$

$$\therefore\ s^2 = \frac{\sum_{i=1}^{n} y_i^2}{n} - \frac{\sum_{i\neq j=1}^{n} y_i y_j}{n(n-1)}$$

Now,

$$E(s^2) = E\left(\frac{\sum_{i=1}^{n} y_i^2}{n}\right) - E\left(\frac{\sum_{i\neq j=1}^{n} y_i y_j}{n(n-1)}\right) \quad \text{---(1)}$$

Consider,

$$E\left(\sum_{i=1}^{n} y_i^2\right) = E\left(\sum_{i=1}^{N} a_i y_i^2\right)$$

$$= E(a_i)\sum_{i=1}^{N} y_i^2$$

$$\therefore\ E\left(\sum_{i=1}^{n} y_i^2\right) = \frac{n}{N}\cdot\sum_{i=1}^{N} y_i^2$$

$\because$ (a_i takes value o
1)

$\therefore$ [Prob. of selecting a u
from sample $= \frac{n}{N}$]

Now,

$$E\left(\sum_{i\neq j=1}^{n} y_i y_j\right) = E\left(\sum_{i\neq j=1}^{N} a_i a_j y_i y_j\right)$$

$$= E(a_i a_j)\sum_{i\neq j=1}^{N} y_i y_j$$

$$= E[0\times 0 \cdot P(0,0) + 1\times 0 \cdot P(1,0) + 0\times 1 \cdot P(0,1) + 1\times 1 \cdot P(1,1)] \sum_{i\neq j=1}^{N} y_i y_j$$

$$E\left(\sum_{i\neq j=1}^{n} y_i y_j\right) = \frac{n}{N}\cdot\frac{(n-1)}{(N-1)}\cdot\sum_{i\neq j=1}^{N} y_i y_j$$

$\therefore$ Eq^n (1) becomes,

$$E(s^2) = \frac{1}{\cancel{n}} \cdot \frac{\cancel{n}}{N} \sum_{i=1}^{N} y_i^2 - \frac{1}{\cancel{n(n-1)}} \cdot \frac{n\cancel{(n-1)}}{N(N-1)} \sum_{i \neq j = 1}^{N} y_i y_j$$

$$\therefore E(s^2) = \frac{\sum_{i=1}^{N} y_i^2}{N} - \frac{\sum_{i \neq j = 1}^{N} y_i y_j}{N(N-1)} \quad \text{—— (2)}$$

We know,

$$\left(\sum_{i=1}^{N} y_i\right)^2 = \sum_{i=1}^{N} y_i^2 + \sum_{i \neq j = 1}^{N} y_i y_j$$

$$\therefore \sum_{i \neq j = 1}^{N} y_i y_j = \left(\sum_{i=1}^{N} y_i\right)^2 - \sum_{i=1}^{N} y_i^2$$

Eq^n(2) becomes,

$$\therefore E(s^2) = \frac{\sum_{i=1}^{N} y_i^2}{N} - \frac{1}{N(N-1)}\left[\left(\sum_{i=1}^{N} y_i\right)^2 - \sum_{i=1}^{N} y_i^2\right]$$

$$= \frac{\sum_{i=1}^{N} y_i^2}{N} - \frac{\left(\sum_{i=1}^{N} y_i\right)^2}{N(N-1)} + \frac{\sum_{i=1}^{N} y_i^2}{N(N-1)}$$

$$= \left(\frac{1}{N} + \frac{1}{N(N-1)}\right) \sum_{i=1}^{N} y_i^2 - \frac{\left(\sum_{i=1}^{N} y_i\right)^2}{N(N-1)}$$

$$= \left\{\frac{N[N-1+1]}{N^2(N-1)}\right\} \sum_{i=1}^{N} y_i^2 - \frac{\left(\sum_{i=1}^{N} y_i\right)^2}{N(N-1)}$$

$$\therefore E(s^2) = \frac{1}{N-1} \sum_{i=1}^{N} y_i^2 - \frac{\left(\sum_{i=1}^{N} y_i\right)^2}{N(N-1)}$$

We know,

$$\bar{y}_N = \frac{\sum_{i=1}^{N} y_i}{N}$$

$$\therefore \sum_{i=1}^{N} y_i = N\bar{y}_N$$

$$\cancel{\therefore E(s^2) = \frac{1}{N-1}(N\bar{y}_N) -}$$

$$E(s^2) = \frac{1}{N-1}\sum_{i=1}^{N} y_i^2 - \frac{N^{\cancel{2}}\bar{Y}_N^2}{\cancel{N}(N-1)}$$

$$= \frac{\sum_{i=1}^{N} y_i^2 - N\bar{Y}_N^2}{N-1}$$

$$E(s^2) = \sum_{i=1}^{N}\frac{(y_i - \bar{Y}_N)^2}{N-1}$$

8 $\therefore \boxed{E(s^2) = S^2}$

Hence, proved.

Q.2] i) • Sampling frame :-

- A list of sampling units is called as 'sampling frame'. It determines structure of sampling survey. Selection of sampling uni & estimation of population parameters are completely dependent on construction of 'sampling frame'.

- Following are important characteristics that a ideal sampling frame should possess :

1) Only relevant units should be listed in sampling frame, otherwise relevant informa would not be possible
2) It should be uptodate. The outdated da should be avoided & regular updation of frame should be done.
3) It should contain entire population i.e. every element of population of intrest should be present in sampling frame.

4) It should be organized in logical & systematic manner.
5) There should not be ambiguity in selecting sampling units or there should not be problem in identifying sampling unit.
6) It should be free from error of duplication ∴ every element of population should be present only once in sample.

3.2] ii) • Sampling: A finite subset of population is called as 'sample'. A process of selecting a sample from population is called as 'sampling'.

• Advantages of sampling over census method:
1) It requires less time to collect & process the data & produce results as compared to census survey, because a part of population is examined.
2) There is reduction in cost in terms of money & man power.
3) Results based on sampling as more accurate & reliable because in sample survey, it may be possible to use better resources.
4) It has better scope. It may include high trained person or specialized equipments.

in some ~~co~~ inquiries, for collection of data.

5) If population is too large, we don't have any alternative.
   e.g., no. of trees in jungle,
   no. of stars in sky.

6) If testing is ~~destructible~~ destructive, it is impracticable to use census method. e.g., cra

7) If population is hypothetical, sampling is th only scientific method.
   e.g., Tossing a coin.

~~iii) Sampling Errors:~~
~~Errors involved in collection, processing & analysis of data are classified into 2 typ~~
~~1) Sampling errors~~
~~2) Non-consuming errors.~~

Q.2] iv) • Population: The group of individuals under study is called as 'population' or 'universe'.

The population may be finite or infinite depending on number of units in it. The population size is no. of individuals in a population & denoted by 'N'.

• Sampling unit: An element or group of elements which are chosen as unit for purpose of enumeration is called as 'sampling unit'.

e.g., 1) No. of students in a class, sampling unit is students.

2) In socio-economic survey for selecting people in town, the sampling units might be an individual person, ~~a~~ family, a household or a block in a city.

3

Krushna C. Petkar. 27-08-24

Page No. 19/20

"ज्ञान, विज्ञान आणि सुसंस्कार यांसाठी शिक्षण प्रसार"
- शिक्षणमहर्षी डॉ. बापूजी साळुंखे

Shri Swami Vivekanand Shikshan Sanstha's

VIVEKANAND COLLEGE, KOLHAPUR.

DEPARTMENT OF MATHEMATICS

Class: _____ Roll No.: _____ Date: _____ Sign _____

Name of the Expt.: _____ W. E. No. _____

Q.5 In SRSWOR, show that, $E(s^2) = S^2$.

The sample mean square is,

$$s^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y}_n)^2}{n-1} = \frac{\sum_{i=1}^{n}(y_i-\bar{y}_n)^2}{n-1}$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n} y_i^2 - n\bar{y}_n^2\right)$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - n\left(\frac{\sum_{i=1}^{n} y_i}{n}\right)^2\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - n\left(\frac{\sum_{i=1}^{n} y_i^2}{n}\right)\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 + \left(-\sum_{i\neq j=1}^{n} y_i y_j + \sum_{i=1}^{n} y_i^2\right)\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 + \left(\frac{\sum_{i=1}^{n} y_i^2 - \sum_{i\neq j=1} y_i y_j}{n}\right)\right]$$

$$= \frac{1}{n-1}\left[\frac{\sum_{i=1}^{n} y_i^2}{n} + \frac{\sum_{i=1}^{n} y_i^2}{n} - \frac{\sum_{i\neq j=1}^{n} y_i y_j}{n}\right]$$

$$= \frac{1}{n-1}\left[\frac{\sum y_i^2 + \sum y_i^2}{n} - \frac{\sum_{i\neq j=1} y_i y_j}{n}\right]$$

$$= \frac{1}{n(n-1)}\left[(n-1)\sum y_i^2 - \frac{\sum_{i\neq j=1}^{n} y_i y_j}{n}\right]$$

$$= \frac{1}{1}$$

$$S^2 = \frac{\Sigma y_i^2}{n} - \sum_{i \neq j = 1} \frac{y_i y_j}{n(n-1)}$$

$$E(S^2) = E\left[\frac{\Sigma y_i^2}{n} - \sum_{i \neq j = 1}^{n} \frac{y_i y_j}{n(n-1)}\right]$$

$$E(S^2) = E\left(\frac{\Sigma y_i^2}{n}\right) - E\left(\sum_{i \neq j = 1}^{n} \frac{y_i y_j}{n(n-1)}\right) \quad -(1)$$

Now,

$$E(\Sigma y_i^2) = E\left(\sum_{i=1}^{n} a_i \, y_i^2\right)$$

$$= \sum_{i=1}^{n} E(a_i) \, y_i^2.$$

$$= \frac{n}{N} \quad \Sigma y_i^2 \quad -(2)$$

a_i: 0, 1.

Similarly,

$$\Sigma y_i y_j = E\left(\sum_{i \neq j = 1}^{n} a_i a_j \; y_i y_j\right)$$

$$= \Sigma E(a_i) \, E(a_j) \, y_i y_j$$

$$= \sum_{i \neq j = 1}^{N} E(a_i a_j) \; y_i y_j \quad -(3)$$

$$= 1 \cdot P(a_i = 1, a_j = 1) + 0 \cdot P(a_i = 0, a_j = 0)$$

$$1 \cdot P(a_i = 1, a_j = 0) \;.\; 0 \times P(a_i = 0, a_j = 1)$$

$$\sum_{i \neq j = 1}^{N} y_i y_j$$

$$= \sum_{i \neq j = 1}^{N} y_i y_j$$

$$= 1 P(a_i = 1, a_j = 1) \cdot 1 P(a_i = 1) \; \Sigma y_i y_j$$

$$\sum_{i \neq j = 1}^{n} y_i y_j = \frac{n}{N} \cdot \frac{(n-1)}{N-1} \cdot \sum_{i \neq j = 1}^{N} y_i y_j \quad -(3)$$

Putting the values of eq^n (2) & (3) in eq^n (1) we get

$$E(S^2) = \frac{n}{N} \cdot \frac{\sum y_i^2}{n} - \frac{n}{N} \; \frac{y_i y_j}{N(N-1)}$$

$$= \frac{n}{N} \cdot \frac{\sum_{i=1}^{N} y_i^2}{N} - \frac{1}{n(n-1)} \times \frac{n(n-1)}{N(N-1)} \sum y_i y_j$$

$$= \sum_{i=1}^{N} \frac{y_i^2}{N} - \frac{\sum_{i \neq j=1}^{N} y_i y_j}{N(N-1)}$$

$$\sum_{i=1}^{N} y_i^2 = \left(\sum_{i=1}^{N} y_i\right)^2 + \sum_{i \neq j=1}^{N} y_i y_j$$

$$\sum_{i=1}^{N} y_i^2 - \left(\sum_{i=1}^{N} y_i\right)^2 = \sum_{i \neq j=1}^{N} y_i y_j$$

$$E(s^2) = \frac{\sum_{i=1}^{N} y_i^2}{N} - \frac{1}{N(N-1)} \left(\left(\sum_{i=1}^{n} y_i\right)^2 - \sum_{i=1}^{n} y_i^2 \right)$$

$$= \frac{\sum_{i=1}^{N} y_i^2}{N} - \frac{\left(\sum_{i=1}^{N} y_i\right)^2}{N(N-1)} + \frac{\sum_{i=1}^{N} y_i^2}{N(N-1)}$$

$$= \frac{1}{N} + \frac{1}{N(N-1)} \sum_{i=1}^{N} y_i^2 - \frac{\left(\sum_{i=1}^{N} y_i\right)^2}{N(N-1)}$$

$$= \frac{N(N-1)+1}{N^2(N-1)} \sum_{i=1}^{N} y_i^2 - \frac{\left(\sum_{i=1}^{N} y_i\right)^2}{\cancel{N(N-1)}\, N(N-1)}$$

$$= \frac{N^2}{N^2(N-1)} \sum_{i=1}^{N} y_i^2 - \frac{\left(\sum_{i=1}^{N} y_i\right)^2}{\cancel{N(N-1)}\, N(N-1)}$$

$$= \frac{\sum_{i=1}^{N} y_i^2}{(N-1)} - \frac{\left(\sum_{i=1}^{N} y_i\right)^2}{(N-1)N}$$

$$= \frac{\sum_{i=1}^{N} y_i^2}{(N-1)} - \frac{\left(\sum y_i\right)^2}{N(N-1)}$$

$$= \frac{\sum_{i=1}^{N} y_i^2}{(N-1)} - \frac{\bar{Y}_N^2}{(N-1)} \qquad \frac{\left(\sum y_i\right)^2}{N} = \bar{Y}_N^2$$

$$E(s^2) = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y}_N)^2}{N-1}$$

$$\therefore \quad S^2 = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y}_N)^2}{N-1}.$$

8

Hence $\boxed{E(s^2) = S^2.}$

Q2. 1) Explain sampling frame.

→ i) The list of sampling units ~~are~~ is called Sampling frame. It determines the structure of sample survey.

ii) Selection of data & planning dependes on the contrustion of sampling frame.

iii) Some are the characteristic for sampli frame,

1) The observations / units should be relev irrelavent units should not be selecte

2) The sampling frame should be in logistic & systamatic manner.

3) There should not be ambiguity in selection of units.

4) The data should be up-to-date. Outd data should be avoided.

5) There should ~~be~~ not be duplication. i a unit is selected at once only repetation is not allowed.

6) If population is distructive, it is impracticable to use census method over sampling

eg - Testing life length of ...

2) Define sampling. Explain advantages of sampling over census method.

→ a) Any finite subset of population is called as ~~sampling~~. sample. The process of selecting sample called as sampling.

b) Advantages of Sampling over census method:

i) Sampling method is not time consuming. Less time is used for selection of units, process of analysis. It decreases cost also.

ii) There is reduction of cost /money in terms of man power.

iii) The results from the sampling method are more reliable & accurate. due to a small part of population is studied.

iv) In terms of large data, we can use only sampling method.

eg- To count the total tress in a jungle.

v) The results get from census method can be more biased as compared to sampling method, because better equipments or specilised man power is used in sampling method.

3

iv) Define population, sampling unit.

→ a) Population:-

i) A group of statistical individual under study is called as population.

ii) It is also called as 'universe'.

iii) Population can be finite or ~~infinite~~ infinite depending on the no. of observations in it.

iv) It is denoted by 'N'.

b) Sampling unit:-

An element or group of element which is choosen as a unit for purpose of enumeration is called as sampling unit.

eg. 1) no. of solar ~~pa~~ systems in Kolhapur.
2) The no. of students in a class.

Vivekanand College Kolhapur (Empowered Autonomous)
B.Sc. III Semester VI
Surprise Test
Paper: Statistical Inference II

Date: 11/02/2025 **Total Marks: 10M**

Name of the Student: Krushna Petkar

8/10

---

1) The most preferred confidence interval for a parameter Θ should be….
   a) with shortest width and largest confidence coefficient
   b) with largest width and largest confidence coefficient
   c) based on sufficient statistics d) both (a) and (b)

2) Given that $P(4.4 \leq \mu \leq 15.7) = 0.90$, Which of the following is correct?
   a) The width of confidence interval is 11.3. b) 4.4 and 15.7 are 90% confidence limits of μ.
   c) Probability that μ lies in the interval (4.4, 15.7) is 0.90 d) All (a) to (c) are true

3) If $X_1, X_2,\dots, X_n$ is a random sample from exponential with parameter θ then interval estimate of θ can be obtained by use of….
   a) Normal distribution b) t-distribution
   c) Chi-square distribution d) F-distribution

4) The NP – Lemma provides best critical region for testing ……hypothesis against …. hypothesis
   a) simple, simple b) simple, composite c) composite, simple d) composite, composite

5) If random variable X has $N(\mu,\sigma^2)$ distribution then which of the following is simple null hypothesis?
   a) $|\mu|=0$ b) $\mu=10$ c) $\sigma^2=16$ d) $\mu=10, \sigma^2=16$

6) Which of the following statement is false?
   a) Probability of rejecting H_0 when H_1 is true is known as type II error.
   b) Neyman Pearson test leads to a Most powerful test.
   c) Probability of rejecting H_0 when H_0 is true is known as type I error.
   d) All the above are true

7) LR-test for testing the equality of variances H_0: $\sigma_1^2 = \sigma_2^2$ against H_1: $\sigma_1^2 \neq \sigma_2^2$ of two normal populations with unknown means is equivalent to a….
   a) Student t-test b) F-test c) Chi-square test d) Z-test

8) The M. P test consist in ………… for fixed α
   a) minimizing β b) maximizing β c) minimizing $(1-\beta)$ d) none of these

9) If statistical test T for testing simple null hypothesis against simple alternative is at least as powerful as any other test then it is known as….
   a) UMP – test b) MP – test
   c) LR – test d) None of them

10) If we increase the confidence level, the confidence interval length is ……
   a) decreases b)) increases
   c) Stays as same d) may increase or decrease depending on data.

****

Vivekanand College Kolhapur (Empowered Autonomous)
B.Sc. III Semester VI
Surprise Test
Paper: Statistical Inference II

Date: 11/02/2025 **Total Marks: 10M**

Name of the Student: Jahnavi mane

9/10

---

1) The most preferred confidence interval for a parameter Θ should be….
   a) with shortest width and largest confidence coefficient
   b) with largest width and largest confidence coefficient
   c) based on sufficient statistics
   d) both (a) and (b)

2) Given that $P(4.4 \le \mu \le 15.7) = 0.90$, Which of the following is correct?
   a) The width of confidence interval is 11.3.
   b) 4.4 and 15.7 are 90% confidence limits of μ.
   c) Probability that μ lies in the interval (4.4, 15.7) is 0.90
   d) All (a) to (c) are true

3) If $X_1, X_2, \ldots, X_n$ is a random sample from exponential with parameter θ then interval estimate of θ can be obtained by use of….
   a) Normal distribution
   b) t-distribution
   c) Chi-square distribution
   d) F-distribution

4) The NP – Lemma provides best critical region for testing ……hypothesis against …. hypothesis
   a) simple, simple b) simple, composite c) composite, simple d) composite, composite

5) If random variable X has $N(\mu,\sigma^2)$ distribution then which of the following is simple null hypothesis?
   a) $|\mu|=0$
   b) $\mu=10$
   c) $\sigma^2=16$
   d) $\mu=10, \sigma^2=16$

6) Which of the following statement is false?
   a) Probability of rejecting H_0 when H_1 is true is known as type II error.
   b) Neyman Pearson test leads to a Most powerful test.
   c) Probability of rejecting H_0 when H_0 is true is known as type I error.
   d) All the above are true

7) LR-test for testing the equality of variances H_0: $\sigma_1^2 = \sigma_2^2$ against H_1: $\sigma_1^2 \ne \sigma_2^2$ of two normal populations with unknown means is equivalent to a….
   a) Student t-test
   b) F-test
   c) Chi-square test
   d) Z-test

8) The M. P test consist in ………… for fixed α
   a) minimizing β b) maximizing β c) minimizing $(1-\beta)$ d) none of these

9) If statistical test T for testing simple null hypothesis against simple alternative is at least as powerful as any other test then it is known as….
   a) UMP – test
   b) MP – test
   c) LR – test
   d) None of them

10) If we increase the confidence level, the confidence interval length is ……
   a) decreases
   b)) increases
   c) Stays as same
   d) may increase or decrease depending on data.

****