PRACTICAL -II

(103ST24-ESTIMATION and 104ST24- SAMPLING THEORY)

M.Sc., STATISTICS First Year

Semester – I, Paper-VI

Lesson Writer

Dr. B.Guravaiah
Assistant Professor
Department of Mathematics & Statistics,
Vignan's Foundations for Science,
Technology & Research (Deemed to be University),
Guntur -Tenali Rd, Vadlamudi,
Andhra Pradesh.

Editor:
Prof. G.V.S.R. Anjaneyulu
Professor of Statistics (Retd.)
Department of Statistics
Acharya Nagarjuna University
.

Director, I/c
Prof. V.VENKATESWARLU
MA.,M.P.S., M.S.W.,M.Phil., Ph.D.

CENTREFORDISTANCEEDUCATION
ACHARAYANAGARJUNAUNIVERSITY
NAGARJUNANAGAR – 522510
Ph:0863-2346222,2346208,
0863-2346259(Study Material)
Website: www.anucde.info
e-mail:anucdedirector@gmail.com

M.Sc., STATISTICS – PRACTICAL -II

First Edition2025

No. of Copies :

© **Acharya Nagarjuna University**

This book is exclusively prepared for the use of students of M.SC.(Statistics) Centre for Distance Education, Acharya Nagarjuna University and this book is meant for limited Circulation only.

Published by:
Prof. V.VENKATESWARLU,
Director, I/C
Centre for Distance Education,
Acharya Nagarjuna University

Printed at:

FOREWORD

Since its establishment in 1976, Acharya Nagarjuna University has been forging ahead in the path of progress and dynamism, offering a variety of courses and research contributions. I am extremely happy that by gaining 'A^{+}' grade from the NAAC in the year 2024, Acharya Nagarjuna University is offering educational opportunities at the UG, PG levels apart from research degrees to students from over 221 affiliated colleges spread over the two districts of Guntur and Prakasam.

The University has also started the Centre for Distance Education in 2003-04 with the aim of taking higher education to the doorstep of all the sectors of the society. The centre will be a great help to those who cannot join in colleges, those who cannot afford the exorbitant fees as regular students, and even to housewives desirous of pursuing higher studies. Acharya Nagarjuna University has started offering B.Sc., B.A., B.B.A., and B.Com courses at the Degree level and M.A., M.Com., M.Sc., M.B.A., and L.L.M., courses at the PG level from the academic year 2003-2004 onwards.

To facilitate easier understanding by students studying through the distance mode, these self-instruction materials have been prepared by eminent and experienced teachers. The lessons have been drafted with great care and expertise in the stipulated time by these teachers. Constructive ideas and scholarly suggestions are welcome from students and teachers involved respectively. Such ideas will be incorporated for the greater efficacy of this distance mode of education. For clarification of doubts and feedback, weekly classes and contact classes will be arranged at the UG and PG levels respectively.

It is my aim that students getting higher education through the Centre for Distance Education should improve their qualification, have better employment opportunities and in turn be part of country's progress. It is my fond desire that in the years to come, the Centre for Distance Education will go from strength to strength in the form of new courses and by catering to larger number of people. My congratulations to all the Directors, Academic Coordinators, Editors and Lesson-writers of the Centre who have helped in these endeavors.

Prof. K. Gangadhara Rao
M.Tech.,Ph.D.,
Vice-Chancellor I/c
Acharya Nagarjuna University

M.Sc.–Statistics

Syllabus SEMESTER-I

106ST24:: PRACTICAL- II

(Papers on 103ST24 and 104ST24)

103ST24 – ESTIMATION

1. Finding of Sufficient estimators
2. Finding of Unbiased estimators and consistent estimators
3. Finding of Efficient estimators and relative efficiency of estimators
4. Problems on Cramer-Rao inequality and MVB estimators and UMVUE
5. Estimation using Minimum chi-square method
6. Estimation using Modified Minimum chi-square method
7. Estimation using Method of Maximum likelihood
8. Estimation using Method of Moments
9. Problems on Confidence Intervals for mean and difference of means
10. Problems on Confidence Intervals for proportions.

104ST24 – SAMPLING THEORY

1. Stratified Sampling
2. Gain in Precision Due to Stratification
3. PPS Sampling
4. Gain Efficiency in Stratified Random Sampling
5. Precision of Systematic Sampling and Stratified Sampling
6. Unbiased Estimator of the Population Mean for Systematic Sampling
7. Yield of Average and Standard Error
8. Estimate the Intra-Class Correlation Co-Efficient
9. Ratio Method of Estimation
10. Regression Method of Estimation.

CONTENTS

Course-106ST24 PRACTICAL-II
(Paper on 103ST24 - Estimation)

PRACTICAL - 1

MAXIMUM LIKEITHOOD FUNCTION ZERO TRUNCATE

Problem :

The following truncate poisson data (truncated 0) estimate the parameter by the method of maximum likelihood estimation.

∂i	1	2	3	4	5	6	7	8	9
fi	22	18	18	11	3	6	3	0	1

Aim :

To estimate the parameters by the method of maximum liklehood estimation (zero truncated poisson distribution)

Procedure :

The probability mass function (pmf) of poisson distribution is

$$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{(1-e^{-\lambda})x!}$$

The likelihood function is given by

$$L = \prod_{i=1}^{n} P(xi)$$

$$= \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x}}{(1-e^{-\lambda})x!}$$

$$= \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x}}{(1-e^{-\lambda})x!}$$

$$= \frac{e^{-n\lambda}\lambda^{\sum xi}}{(1-e^{-\lambda})^{n}\prod_{i=1}^{n} xi!}$$

apply log on both sides

$$\log L = \log\left[\frac{e^{-n\lambda}\lambda^{\sum xi}}{(1-e^{-\lambda})^{n}\prod_{i=1}^{n} xi!}\right]$$

$$= \log\left[e^{-n\lambda}\lambda^{\sum xi}\right] - \log\left[(1-e^{-\lambda})^{n}\prod_{i=1}^{n} xi!\right]$$

$$= \left[\log e^{-n\lambda} + \log \lambda^{\sum xi}\right] - \left[\log(1-e^{-\lambda})^n + \log \prod_{i=1}^{n} xi!\right]$$

$$= \log e^{-n\lambda} + \log \lambda^{\sum xi} - \log(1-e^{-\lambda})^n - \log \prod_{i=1}^{n} xi!$$

$$= -n\lambda + \sum_{i=1}^{n} xi \log \lambda - n\log(1-e^{-\lambda}) - \log \prod_{i=1}^{n} xi!$$

$$\frac{\partial \log L}{\partial \lambda} = -n + \frac{\sum xi}{\lambda} - \frac{n}{(1-e^{-\lambda})}(e^{-\lambda}) = 0$$

$$= -n + \frac{\sum xi}{\lambda} - \frac{ne^{-\lambda}}{1-e^{-\lambda}}$$

$$= -n\left[-1 + \frac{\bar{x}}{\lambda} - \frac{e^{-\lambda}}{1-e^{-\lambda}}\right]$$

$$\frac{\partial \log L}{\partial \lambda} = 0$$

$$\Rightarrow n\left[-1 + \frac{\bar{x}}{\lambda} - \frac{e^{-\lambda}}{1-e^{-\lambda}}\right] = 0$$

$$\Rightarrow -1 + \frac{\bar{x}}{\lambda} - \frac{e^{-\lambda}}{1-e^{-\lambda}} = 0$$

$$\Rightarrow \frac{\bar{x}}{\lambda} = 1 + \frac{e^{-\lambda}}{1-e^{-\lambda}}$$

$$\Rightarrow \frac{\bar{x}}{\lambda} = \frac{1-e^{-\lambda}+e^{-\lambda}}{1-e^{-\lambda}}$$

$$\Rightarrow \frac{\bar{x}}{\lambda} = \frac{1}{1-e^{-\lambda}}$$

$$\Rightarrow \bar{x} = \frac{\lambda}{1-e^{-\lambda}}$$

$$\lambda_{i+1} = \bar{x}(1-e^{-\lambda i})$$

using method of itration

$\lambda_{i+1} = \bar{x}(1-e^{-\lambda i})$, i = i, 2........................∞

take $\lambda_0 = 0, \lambda_0 = \bar{x}$

When $|\lambda_{i+1} = \lambda| < 0.01$ stop-the itration until λ_{i+1} as thet estimation of λ_i otherwise the procedure.

Calculations :

First we calculated the mean of the given data.

$$\text{Mean} = \frac{\sum fixi}{N}$$

xi	fi	fixi
1	22	22
2	18	36
3	18	54
4	11	44
5	3	15
6	6	36
7	3	21
8	0	0
9	1	9
	N=82	237

Mean = $\frac{\sum fixi}{N} = \frac{237}{82} = 2.8902$

$\therefore \lambda_{i+1} = \bar{x}\left(1 - e^{-\lambda i}\right)$

let i = 0, take $\lambda = x \equiv 0.28902$

$\lambda_1 = 2.8902(1 - e^{-2.8902})$
$= 2.7297$

let i = 1,

$\lambda_2 = (2.8902)(1 - e^{-2.7297})$
$= 2.7016$

let i = 2,

$\lambda_3 = (2.8902)(1 - e^{-\lambda 2}) = (2.8902)(1 - e^{-2.7016})$
$= 2.6962$

let i = 3,

$\lambda_4 = (2.8902)(1 - e^{-2.6962})$
$= 2.6952$

let i = 4,

$\lambda_5 = (2.8902)(1 - e^{-2.6952})$
$= 2.6950$

let i = 5,

$\lambda_6 = (2.8902)(1 - e^{-2.6950})$
$= 2.6950$

Inference :

The maximum lilkelihood estimation of zero truncated poisson distribution of λ is λ = 2.6949

PRACTICAL - 2

BAYE'S ESTIMATION IS EXPONENTIAL DISTRIBUTION

Problem :

The following data is a sample of 20 deservations from the 1 parameter exponential distribution with mean $\theta = \mu l$

3.4177	2.4651	0.3484	9.2817
2.6155	3.1356	0.3994	1.7823
0.5623	1.8474	2.7425	0.5973
9.3208	3.2257	2.9487	3.5680
3.5251	2.3467	3.7284	1.72271

Assuming the prior density of θ as

$$g\left(\frac{\theta}{x}\right) = \frac{cxp[-a/\theta]}{\theta^c}, c \geq 0, a \geq 0, \theta > 0$$

Obtain the baye's estimator of θ for different values of 'a' and 'c' and comment on the distribution of boye's estimator taking as C = 0, 1, 2 ....

Aim :

To obtain the boye's estimator of θ is, parameter is exponential distribution for the given data and comment on the distribution of the baye's estimator.

Procedure :

The probability density function of exponential distribution is

$$f\left(\frac{x}{\theta}\right) = \frac{1}{\theta}\exp\left[\frac{-x}{\theta}\right], \theta > 0, x > 0$$

θ is treated as random variables and its probablity density function is given by

$$g\left(\frac{\theta}{x}\right) = \frac{g\left(\frac{\theta}{x}\right)\exp\left(\frac{-a}{\theta}\right)}{\theta^c}, c \geq 0, a \geq 0, \theta > 0$$

The baye's estimator of θ is given by $\theta = \frac{n\bar{x}+a}{n+c-2}$

C^a	0	1	2	3	$\frac{Min}{Max}$
0	3.3075	3.3631	3.4186	3.4742	0.9520
1	3.1334	3.1861	3.2387	3.2913	0.9520
2	2.9768	3.0268	3.0768	3.1268	0.9520
3	2.8350	2.8826	2.9302	2.9779	0.9520
$\frac{Min}{Max}$	0.85771	0.8571	0.8571	0.8571	

$$\theta = \frac{n\bar{x} + a}{n + c - 2};\ n = 20;\ \bar{x} = 2.9768$$

$$a = 0,\ c = 0 \Rightarrow \theta = \frac{20(2.9768) + 0}{20 + 0 - 2} = 3.3075$$

$$a = 0,\ c = 1 \Rightarrow \theta = \frac{20(2.9768) + 0}{20 + 1 - 2} = 3.1334$$

$$a = 0,\ c = 2 \Rightarrow \theta = \frac{20(2.9768) + 0}{20 + 2 - 2} = 3.9768$$

$$a = 1,\ c = 0 \Rightarrow \theta = \frac{20(2.9768) + 1}{20 + 0 - 2} = 3.3631$$

$$a = 1,\ c = 1 \Rightarrow \theta = \frac{20(2.9768) + 1}{20 + 1 - 2} = 3.1861$$

$$a = 1,\ c = 2 \Rightarrow \theta = \frac{20(2.9768) + 1}{20 + 2 - 2} = 3.0268$$

$$a = 2,\ c = 0 \Rightarrow \theta = \frac{20(2.9768) + 2}{20 + 0 - 2} = 3.4186$$

$$a = 2,\ c = 1 \Rightarrow \theta = \frac{20(2.9768) + 2}{20 + 1 - 2} = 3.2387$$

$$a = 2,\ c = 2 \Rightarrow \theta = \frac{20(2.9768) + 2}{20 + 2 - 2} = 2.9302$$

$$a = 3,\ c = 0 \Rightarrow \theta = \frac{20(2.9768) + 3}{20 + 0 - 2} = 3.4742$$

$$a = 3,\ c = 1 \Rightarrow \theta = \frac{20(2.9768) + 3}{20 + 1 - 2} = 3.2913$$

$$a = 3,\ c = 2 \Rightarrow \theta = \frac{20(2.9768) + 3}{20 + 2 - 2} = 3.1268$$

$$a = 3,\ c = 3 \Rightarrow \theta = \frac{20(2.9768) + 3}{20 + 3 - 2} = 2.9779$$

Inference :

The baye's estimator for the given data are more rudestress to changes in A & C values.

PRACTICAL - 3

METHOD OF MAXIMUM CHI-SQUARE DISTRIBUTION

Problem :

Type	fi	Probability
long and purple	296	$\frac{2+\theta}{4}$
long and red	27	$\frac{1-\theta}{4}$
round and purple	19	$\frac{1-\theta}{4}$
round and red	85	$\frac{\theta}{4}$

From above data estimate θ value by using $\lambda^2 - dist^n$ where to estimate by the method of λ^2 data by the estimate.

Procedure :

Maximum likelihood equation is given by λ^2.

$$\lambda^2 = \frac{\sum_{i=1}^{n}(Npi - fi)^2}{fi}$$

The maximum likelihood equation to be θ then find the solve for the estimating θ is $\frac{d\lambda^2}{d\theta} = 0$.

If $\theta^{\wedge}$ is the ML estimator of the θ then find teh expected values of θ,

the calculate $x^2 = \sum \frac{\theta_i^{\,2}}{ei} - N$

Conclusion :

If calculate the x^2 value less than or equal to tabulated x^2 value. We accept the null hypothesis otherwise we reject the null hypothesis.

$$x^2 = \sum_{i=1}^{n}(Npi - fi)^2$$

$$x^2 = \frac{(Npi - fi)^2}{fi} + \frac{(Np_2 - f_2)^2}{f_2} + \frac{(Np_3 - f_3)^2}{f_3} + \frac{(Np_4 - f_4)^2}{f_4}$$

$$= \frac{\left[427\left(\frac{2+\theta}{4}\right) - 296\right]^2}{296} + \frac{\left[427\left(\frac{1-\theta}{4}\right) - 27\right]^2}{27} + \frac{\left[427\left(\frac{1-\theta}{4}\right) - 19\right]^2}{19} + \frac{\left[427\left(\frac{\theta}{4}\right) - 85\right]^2}{85}$$

$$\frac{dx^2}{d\theta} = \frac{(427)^2}{296}\left[\left(\frac{2+\theta}{4}\right) - \left(\frac{296}{427}\right)\right]^2 + \frac{(427)^2}{27}\left[\left(\frac{1-\theta}{4}\right) - \left(\frac{27}{427}\right)\right]^2 +$$

$$\frac{(427)^2}{296}\left[\left(\frac{1-\theta}{4}\right) - \left(\frac{19}{427}\right)\right]^2 + \frac{(427)^2}{85}\left[\left(\frac{\theta}{4}\right) - \left(\frac{85}{427}\right)\right]^2$$

$$\Rightarrow \frac{(427)^2}{19}\left[2\left(\frac{2+\theta}{4}\right) - \left(\frac{296}{427}\right)\left(\frac{1}{4/2}\right)\right] + \frac{(427)^2}{27}\left[2\left(\frac{1-\theta}{4}\right) - \left(\frac{27}{427}\right)\left(\frac{-1}{4/2}\right)\right]$$

$$+ \frac{(427)^2}{19}\left[2\left(\frac{1-\theta}{4}\right) - \left(\frac{19}{427}\right)\left(\frac{-1}{4/2}\right)\right] + \frac{(427)^2}{85}\left[2\left(\frac{\theta}{4}\right) - \left(\frac{85}{427}\right)\left(\frac{1}{4/2}\right)\right]$$

$$\Rightarrow \frac{(427)^2}{2\times 296}\left[\left(\frac{2+\theta}{4}\right) - 0.6932\right] - \frac{(427)^2}{2\times 27}\left[\left(\frac{1-\theta}{4}\right) - 0.0632\right]$$

$$- \frac{(427)^2}{2\times 19}\left[\left(\frac{1-\theta}{4}\right) - 0.0444\right] + \frac{(427)^2}{2\times 85}\left[\left(\frac{\theta}{4}\right) - 0.1990\right]$$

$$\Rightarrow 307.988\left[\left(\frac{1}{2} + \frac{\theta}{4}\right) - 0.6932\right] - 3376.46\left[\left(\frac{1}{4} - \frac{\theta}{4}\right) - 0.063\right]$$

$$-4798.13\left[\left(\frac{1}{4} - \frac{\theta}{4}\right) - 0.444\right] + 1072.52\left[\frac{\theta}{4} - 0.199\right]$$

$$\Rightarrow (153.994 + 76.997\theta - 213.49) - (844.1150 + 844.1550 - 212.71698)$$

$$-1199.53 + 119.53\theta + 213.036972 + 268.13\theta - 213.431$$

$$\Rightarrow 76.997\theta + 844.155\theta + 1199.53\theta + 268.13\theta - 1890.819$$

$$\Rightarrow 2388.812\theta - 1890.819$$

$$\theta = \frac{1890.819}{2388.812}$$

$$\theta = 0.7915$$

Observed frequency θi	Probability values	Expected freq $ei = nPi$	$\frac{\theta i^2}{ei}$
296	$\frac{2+\theta}{4} = P_1$	297.9	294.0033
27	$\frac{1-\theta}{4} = P_2$	22.25	32.764
19	$\frac{1-\theta}{4} = P_3$	22.25	16.224
85	$\frac{\theta}{4} = P_4$	84.49	85.51
427		426.89 N427	1128.5043

$$x^2 = \frac{\sum oi^2}{ci} - N$$

$$= 428.5043 - 427$$

$$= 1.5043$$

Here the table vlaue of x^2 at 5% I.O.S. is 5.99.

Here the calculated x^2 value lessthan the table value of x^2. So, we conclude that for the given data the estimated parameter value is good estimator.

Inference :

For the given data we estimated parameter value by the method of maximum x^2 estimation is good estimator.

PRACTICAL - 4

METHOD OF MODIFIED MINIMUM X^2

Problem :

In a genatic problem are Tea plants the following data is obtain.

Type	Observed for frequency	Theritical
long and purple	336	$\frac{2+\theta}{4}$
long and red	57	$\frac{1-\theta}{4}$
round and purple	54	$\frac{1-\theta}{4}$
round and red	105	$\frac{\theta}{4}$
	552	

Estimate the parameter θ by the method of modifies minimum x^2 and test for goodness of fit.

Aim :

To estimate the parameter θ by the method of modified minimum x^2 and test for goodness of fit.

procedure :

The method of modified minimum x^2 statistic to estimate the parameter using the given data is given by $x^2 = \sum_{i=1}^{n}$

$$x^2 = \sum_{i=1}^{n} \frac{(oi-ei)^2}{ei}$$

$$= \sum_{i=1}^{n} \frac{}{oi} \longrightarrow ①$$

Where, oi is the observed frequency of the ith class "ei" is the expected frequency of the ith class

$ei = nPi(\theta)$, where N = total frequency

To estimate the parameter θ we differentiate eq^n ① w.r.t. θ and equating 0 which on will be estimate value of 'θ'.

Here if we necessary to use the iterative method to find the estimate value of 'θ' which ever the expected is 0 which is replaced by 1.

Test for goodness of fit : To test the goodness of fit of the theritical model the null hypothesis (Ho) is the theritical model is good fit for the given data.

The test statistic is given by

$x^2 = \sum_{i=1}^{n} \frac{(oi-ei)^2}{ei} Nf \; x^2_{(n-1)}$ degree of freedom at the given level of significance at 5% C.D.S.

Conclusion :

If the calculated value less than the tabulated value we accept the null hypothesis otherwise we rejected.

Calculations :

$$x^2 = \sum \frac{(NPi)^2}{fi} - N$$

$$x^2 = \left[\frac{(NPi)^2}{fi} + \frac{(NP_2)^2}{f_2} + \frac{(NP_3)^2}{f_3} + \frac{(NP_4)^2}{f_4}\right]$$

$$= \left[\frac{\left[552\left(\frac{2+\theta}{4}\right)\right]^2}{336} + \frac{\left[552\left(\frac{1-\theta}{4}\right)\right]^2}{57} + \frac{\left[552\left(\frac{1-\theta}{4}\right)\right]^2}{54} + \frac{\left[552\left(\frac{\theta}{4}\right)\right]^2}{105}\right] - 552$$

$$= \left(\frac{552}{4}\right)^2\left[\frac{(2+\theta)^2}{336} + \frac{(1-\theta)^2}{57} + \frac{(1-\theta)^2}{54} + \frac{(\theta/4)^2}{105}\right] - 552$$

$$\frac{\partial x^2}{\partial\theta} = \left(\frac{552}{4}\right)^2\left[2\left(\frac{2+\theta}{336}\right) + \frac{2(1-\theta)}{57}(-1) + \frac{2(1-\theta)}{54}(-1) + \frac{2\theta}{105}\right]$$

Now, $\frac{\partial x^2}{\partial\theta} = 0$

$$\Rightarrow \left(\frac{552}{4}\right)^2\left[\frac{2(2+\theta)}{336} + \frac{2(\theta-1)}{57} + \frac{2(\theta-1)}{54} + \frac{2\theta}{105}\right] = 0$$

$$\Rightarrow \frac{4}{36} + \frac{2\theta}{336} + \frac{2\theta}{57} - \frac{2}{57} + \frac{2\theta}{54} - \frac{2}{54} + \frac{2\theta}{105} = 0$$

$$\Rightarrow \frac{4}{336} + \frac{2}{57} - \frac{2}{54} + \theta\left[\frac{2}{336} + \frac{2}{57} + \frac{2}{54} + \frac{2}{105}\right] = 0$$

$$\Rightarrow 0.0119 - 0.0350 - 0.0370 + \theta\left[0.0059 + 0.0350 + 0.0370 + 0.0190\right] = 0$$

$$\Rightarrow -0.0602 + (0.0969)\theta = 0$$

$$\theta = \frac{0.0602}{0.0969}$$

$$\theta = 0.6213$$

Now the theritical probabilities is

$$\Pi_1(\theta) = \frac{2+\theta}{4} = \frac{2+0.6213}{4} = 0.65532$$

$$\Pi_2(\theta) = \frac{1-\theta}{4} = \frac{1-0.6213}{4} = 0.09467$$

$$\Pi_3(\theta) = \frac{\theta}{4} = \frac{0.6213}{4} = 0.15532$$

$$\Pi_4(\theta) = \frac{\theta}{4} = \frac{0.6213}{4} = 0.15532$$

$\Pi_i(\theta)$	$N(\Pi_i\theta) = ei$	θ_i	$(oi - ei)$	$(oi-ei)^2$	$\frac{oi-ei}{ei}$
0.6553	361.7256	336	-25.7256	661.8064	1.8295
0.0946	52.2192	57	4.7804	22.8522	0.4375
0.0946	52.2192	54	1.7808	3.17122	0.0607
0.1553	85.7256	105	19.2744	371.5024	4.3336
		551.88 N552	552		1059.3360

$$x^2 = \frac{\sum(oi-ei)^2}{ci} = 6.6614$$

Inference :

Using the method of modified minimum x^2 the estimated value of parameter θ is $\theta^n = 0.6213\cdot$

Here the calculated x^2 value lessthan the table value of the x^2 with x^2_{n-1} degree of fredom at 5% of I.O.S.

So we accept the null hypothesis and we conclude that the theritical model is good fit for the given problem.

PRACTICAL - 5

FINDING SUFFICIENT ESTIMATORS

Problem : Coin flips (Bernouli trials)

Suppose we flip a coin n = 3 times, with probability 'P' of heads (1) and 1-P of tails (0).

We observe two data sets Data

set : A : (1, 0, 1)

Data set : B : (1, 1, 0)

Both data sets have same sufficient statistic : The

total no. of heads T(X) = 2.

Aim :

To find out the both data sets have the same sufficient statistic (or) not.

Procedure (&) Calculation :

Here we can findout the sufficient statistic we follow the given steps. Step 1 :

Joint probability man function (PMF)

For any dataset T = t, the joint P.M.F. is

$$f(x/p) = P^t.(1-P)^{n-t}$$

For both data sets

$$f(x/p) = P^t.(1-P)^1$$

Step 2 : Factorization theorem

Using the Fisher - Ney man factorization theorem

$$f(x/p) = P^t.(1-P)^1.1$$

$$g(T(x), P) \quad h(x)$$

Hence $g(T(x),P) = P^{T(x)}(1-P)^{n-T(x)}$ depends on the data only through $T(x)$ and $h(x)=1$ is Independence of 'P'. Thus $T(x)=\sum xi$ is sufficient.

Step 3 : Likelihood comparison

For P = 0.5 and P = 0.6

Data set A :

$$L(P=0.5) = 0.5^2 \times 0.5 = 0.125,$$

$$L(P=0.6) = 0.6^2 \times 0.4 = 0.144$$

Data Set B :

$$L(P=0.5)=0.5^2\times 0.5=0.125,$$

$$L(P=0.6)=0.6^2\times 0.4=0.144$$

The likelihood x are identical because they depend only on T = 2, not the data order.

Step 4 : Conditional distribution

Given T = 2 there are $\binom{3}{2}$ = 3 possible sequencies.

The conditional distribution of the given data T = 2 is Uniform 1/3 for each sequence 1, independent of P. This is comfirm sufficiency.

Conclusion :

The total no. of heads $T(x)$ is sufficient statistic for P. Once $T(x)$=2 is known the specific sequence of outcomes provides no additional information about P.

$T(x)=\sum_{i=1}^{n} x_i$ is sufficient estimator of ‘P.

PRACTICAL - 6

FINDING COMBINED ESTIMATOR OF CONSISTENT ESTIMATOR

Problem :

Suppose a true population mean= 10 and we take random samples of different sizes to estimate ' ' using the sample mean.

We will 1. Check unbiasedness of $\mu^{\wedge}$ for small sample.

2. Check conistency by increasing 'n' and observing has $\mu^{\wedge}$ behaves.

Aim :

To Find the unbiased estimator of constituency estimator.

Procedure (or) Calculation :

For small sample (n = 3)

Suppose we randomly sample three values from a population with mean $\mu^{\wedge}$ =10.

$x_1 = 8,\ x_2 = 12,\ x_3 = 10$

Compute sample mean $\mu^{\wedge} = \dfrac{8+12+10}{3} = \dfrac{30}{3} = 3$

Check unbiasedness

To check if is unbiased, compute $E(\hat{\mu})$

$$E(\hat{\mu}) = E\left[\frac{x_1 + x_2 + x_3}{3}\right] = \frac{E(x_1) + E(x_2) + E(x_3)}{3}$$

Since $E(x_i) = \mu = 10$ we get $E(\mu^{\wedge}) = \dfrac{3(10)}{3} = 10$

Since $E(\mu^{\wedge}) = \mu^{\wedge}$ the sample mean is unbiased estimator of ' '.

For large sample : (n = 10) to check conisstency Now we take the large sample.

x = [9, 10, 11, 8, 12, 9, 10, 11, 10, 10]

Compute sample mean

$$\mu^{\wedge} = \frac{9+10+11+8+12+9+10+11+10+10}{10}$$

$$\mu^{\wedge} = \frac{100}{10} = 10$$

As 'n' increases the sample mean stays close to 10, including that $\mu^{\wedge}$ is consistent estimator of ' '.

Inference :

Unbiasedness : The expected value of $\mu^{\wedge}$ is proving it's an onbiased estimator.

Consistuency : As we take large samples $(n \to \infty), (\mu^{\wedge} \to \mu)$, confirming consistency.

PRACTICAL- 7

FINDING EFFICIENCY ESTIMATOR AND RELATIVE EFFICIENCY ESTIMATOR

Problem :

A reasearcher is estimating population mean ‘ ’ using two different estimates.

1. Estimator 1 : $\hat{\mu}_1$ with variance var$(\hat{\mu}_1)$ = 2.
2. Estimator 2 : $\hat{\mu}_2$ with variance var $(\hat{\mu}_2)$ = 5.

Aim :

To find out the efficiency estimator and relative efficiency estimator.

Procedure and calcuations :

The more efficient estimator is the one with the smaller variance.

The relative efficiency of $\hat{\mu}_1$ compared $\hat{\mu}_2$ is given by

$$RE(\hat{\mu}_1, \hat{\mu}_2) = \frac{Var(\hat{\mu}_1)}{Var(\hat{\mu}_2)}$$

Step 1 :

Determine the more efficient estimator :

The efficiency of an estimator is determined by it’s variance.

The smaller variance, the more efficient estimator.

Since $Var(\hat{\mu}_1)$ = 2 is smaller than

$Var(\hat{\mu}_2)$= 5, is $\hat{\mu}_1$ is more efficient estimator

Step 2 :

Compute Relative Efficiency :

To relative efficiency of $\hat{\mu}_1$ compared to $\hat{\mu}_2$ is given by

$$RE(\hat{\mu}_1, \hat{\mu}_2) = \frac{Var(\hat{\mu}_1)}{Var(\hat{\mu}_2)}$$

Substitute the values

$$RE(\hat{\mu}_1, \hat{\mu}_2) = \frac{5}{2} = 2.5$$

Relative Efficiency is “2.5”.

Inference :

Since RE = 2.5, it means that $\hat{\mu}_1$ is 2.5 times more efficient than $\hat{\mu}_2$. This means that $\hat{\mu}_1$ provides a more precise estimate of with less variability.

PRACTICAL - 8

PROBLEM ON CR-RAO INEQUALITY AND MVB ESTIMATE UMVUE

Problem :

We have a random sample of size n = 6 from an exponential distribution with mean 'θ'.

$x_1, x_2, \ldots\ldots\ldots x_n \sim E_{xp}(\theta)$. We observe the following data $x_1 = 2, x_2 = 3, x_3 = 1, x_4 = 4, x_5 = 2, x_6 = 5$.

The p.d.f (Probability density function) is

$$f(x:\theta) = \frac{1}{\theta}.e^{-x/\theta}, x > 0$$

Using above data. Solve the following.

Compute the CR-Rao inequality and MVB estimate UMVUE.

Aim :

To compute the CR-Rao inequality and MVB estimate UMVUE.

Procedure (&) Calculations :

Compute Cramer-Rao lower bound

For an exponential distribution, the fisher information is $I(\theta) = \frac{n}{\theta^2}$

Thus, the Cramer-Rao lower bound (CRLB) for the variance of any onbiased estimator $\theta\hat{}$ is

$$Var(\hat{\theta}) \geq \frac{1}{I(\theta)} = \frac{\theta^2}{n}$$

Find an onbiased estimator 'θ'.

For an exponential distribution, the sample mean

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} xi$$

is an onbaised estimator of 'θ' because

$$E(\bar{x}) = \theta$$

Since the sample mean is function of the sufficient statistic $T = \sum xi$ and is onbaised, by Lehmann-sheffer's theorem, $\bar{x}$ is the UMVUE of 'θ'.

This we take

$$\hat{\theta} = \bar{x}$$

Estimating 'θ' using the given data

First we compute the sample mean

$$\bar{x} = \frac{2+3+1+4+2+5}{6}$$

$$= \frac{17}{6} \simeq 2.83$$

Check if $\hat{\theta}$ acheives the CRLB.

We know that the variance of the sample mean for an exponential distribution is

$$Var(\bar{x}) = \frac{\theta^2}{n}$$

Which matches the CRLB $Var(\hat{\theta}) = \dfrac{\theta^2}{n}$

Since the variance of $\hat{\theta}$ meets the lower band it is an efficient estimator meaning that it achieves to Cramer-Rao bound.

PRACTICAL - 9

ESTIMATION USING METHOD OF MOMENTS

Problem : Suppose $x_1, x_2, ..., x_n$ is a random sample from a Beta distribution with parameters α and β.

$x \sim Beta(\alpha, \beta)$

The p.d.f. is $f(x, \alpha, \beta) = \dfrac{x^{\alpha-1} e^{-x/\beta}}{\beta^{\alpha} \Gamma(\alpha)}$

We observe ethe following data

$x_1 = 3, x_2 = 4, x_3 = 2, x_4 = 5, x_5 = 6$

Using the method of moments estimate 'α' and 'β'.

Aim :

To estimate α, β using the method of moments.

Procedure (&) Calculation :

Compute population moments

For a Gamma (α, β) distribution

$E(x) = \alpha\beta$

$Var(x) = \alpha\beta^2$

Compute sample moments

Sample mean $\bar{x} = \dfrac{3+4+2+5+6}{5} = \dfrac{20}{5} = 4$

Sample variance : $S^2 = \dfrac{1}{n-1} \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2$

$= \frac{1}{4}\left[(3-4)^2 + (4-4)^2 + (2-4)^2 + (5-4)^2 + (6-4)^2\right]$

$= \dfrac{1}{4}[1+0+4+1+4]$

$= \dfrac{10}{4} = 2.5$

Solve α and β

From the moments equations

1. $\alpha\beta = \bar{x} = 4 \longrightarrow (1)$
2. $\alpha\beta^2 = S^2 = 2.5 \longrightarrow (2)$

Dividing equation (2) by equation (1)

$$\frac{\alpha/\beta^2}{\alpha/\beta} = \frac{2.5}{4}$$

$$\beta = \frac{2.5}{4} = 0.625$$

Substituting the β into equation '1'

$$\alpha\,(0.625) = 4$$

$$\alpha = \frac{4}{0.625}$$

$$\alpha = 6.4$$

Using the method of moments, we estimates

$$\hat{\alpha} = 6.4,\ \hat{\beta} = 0.625$$

Inference :

The find estimates are

$$\hat{\alpha} = 6.4 \quad and \quad \hat{\beta} = 0.625$$

PRACTICAL - 10

PROBLEMS ON CONFIDENCE INTERVAL FOR PROPORTIONS

Problem :

A company wants to estimate the promotion of customers who prefer their new product over the old one. A random sample of 500 customers is surveyed and 320 of them say they prefer the new product. Construct a 95% confidence interval for the true proportion of customers who prefer the new product.

Aim :

To find out confidence interval for proportions.

Procedure (&) Calculations :

Given information

Sample size n = 500

Sample proportion = $\hat{p} = \frac{x}{n} = \frac{320}{500} = 0.64$

Confidence level : 95%

For a proportion, the confidence interval is given by

$$\hat{p} = Z_{x/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Where $Z_{a/2}$ is the critical lvalue from the standard normal table for 95% confidence level.

$$Z_{0.025} = 1.96$$

'$\hat{p}$' is the sample proportion

'n' is the sample size.

Compute Standard Error :

$$S.E = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$= \sqrt{\frac{0.64(1-0.64)}{500}}$$

$$= \sqrt{\frac{0.64 \times 0.36}{500}}$$

$$= \sqrt{\frac{0.2304}{500}}$$

$$= \sqrt{0.0004608} = 0.0215$$

Compute the margin of Error (M.E) :

$$M.E = Z_{\alpha/2} \times S.E$$

$$= 1.96 \times 0.0215$$

$$= 0.0421$$

Compute the Confidence Interval :

$$= \hat{p} \pm ME$$

$$= 0.64 \pm 0.0421$$

$$= (0.5979, 0.6821)$$

Thus the 95% confidence interval for the true proportion of customers who prefer the new product is (0.598, 0.682)

Inference :

We are 95% confident that the treu proportion of customers who prefer the new product is b/w 59.8% and 68.2%.

Course-106ST24 PRACTICAL-II
(Paper on 104ST24 - Sampling Theory)

PRACTICAL - 1

STRATIFIED RANDOM SAMPLING

Problem :

The following data shows the stratification of all the forms, in country by farm size and the average acres of corn (maize) per farm in each strataum. For a sample of 100 farms complete the sample sizes in each stratum under

a) Proportional allocation

b) Neyman allocation

Compare the precisions of these methods with that of SRS where FPC is ignored.

Farm Size	Number of farms (Nh)	Average corn acres (YNh)	Standard Deviation
0 - 40	394	5.4	8.3
41 - 80	461	16.3	13.3
81 - 120	391	24.3	15.1
121 - 160	334	34.5	19.8
161 - 200	169	42.1	24.5
201 - 240	113	50.1	26.0
241	148	63.8	35.2

Aim :

1) To compute the sample size under proportional allocation and Neyman alloca- tions.

2) To compute the variance due to

a) Proportional allocation

b) Neyman allocation

3) Also compare the precision of these methods with the simple random sampling when FPC is ignored.

Procedure :

When FPC is ignored

$$V(prop) = \sum_h \frac{N_h S_h^2}{nN}$$

$$V(ney) = \frac{\left(\sum_h N_h S_h\right)^2}{nN^2}$$

$$Vran = \frac{1}{n}\frac{\left(\sum_h (N_h - 1)S_h^{\ 2} + \sum_h N_h (Y_h - Y)^2\right)}{(N-1)}$$

For relative precision for proportion allocation.

$$= \frac{\frac{1}{V_{prop}}}{\frac{1}{V_{ran}}} \times 100$$

For relative precision for Neyman allocation.

$$= \frac{\frac{1}{V_{ney}}}{\frac{1}{V_{ran}}} \times 100$$

Proportional allocation $n_h = \left(\frac{n}{N}\right) N_h$

Neyman allocation $n_h = \frac{n}{\left(\sum_h N_h S_h\right)} . N_h . S_h$

Calculations :

Farm Size	N_h	$\overline{Y}_{N_h}$	S_h	S_h^2	$N_h S_h$	$N_h S_h^2$	$(N_h S_h)^2$	$N_h.\overline{Y}_{N_h}$	$(\overline{Y}_{N_h} - \overline{Y})^2$	$N_h(\overline{Y}_{N_h} - \overline{Y})^2$	$(N_h - 1)S_h^2$
0-40	394	5.4	8.3	68.89	3270.2	27142.66	10694208.04	2127.6	436.81	172103.14	27073.77
41-80	461	16.3	13.3	176.89	6131.3	81546.29	37592839.69	7514.3	100	46100	81369.4
81-120	391	24.3	15.1	228.01	5904.1	89151.91	34858396.81	9501.3	4	1564	88923.9
121-160	334	34.5	19.8	392.04	6613.2	130941.36	43734414.24	11523	67.24	22458.16	130549.32
161-200	169	42.1	24.5	600.25	4140.5	101442.25	17143740.25	7114.9	249.64	42189.16	100842
201-240	113	50.1	26.0	676	2938	76388	8631844	5661.3	566.44	64007.72	75712
241	148	63.8	35.2	1239.04	5209.6	183377.92	27139932.16	9442.4	1406.25	208125	182138.88
	$\sum N_h =$ 2010			$\sum S_h^2 =$ 3381.12	$\sum N_h S_h =$ 34206.9	$\sum N_h S_h^2 =$ 689990.39	$(\sum N_h S_h)^2 =$ 179795375.2	$\sum N_h \overline{Y}_{Nh} =$ 52884.8	$\sum(\overline{Y}_{Nh} - \overline{Y})^2 =$ 2830.38	$\sum N_h(\overline{Y}_{N_h} - \overline{Y})^2 =$ 556547.18	$\sum(N_h - 1)S_h^2 =$ 686609.27

$$Y = \frac{\sum N_h \overline{Y}_{N_h}}{\sum N_h} = \frac{52884.8}{2010} = 26.3$$

$$V_{prop} = \frac{1}{nN} \cdot \sum N_h . S_h^{\ 2}$$

$$= \frac{1}{(100)(2010)} (689990.39)$$

$$V_{prop} = 3.4327$$

$$V_{ney} = \frac{1}{nN^2} \left(\sum N_h S_h \right)^2$$

$$= \frac{1}{(100)(2010)^2} (179795375.2)$$

$$V_{ney} = 2.8962$$

$$V_{ran} = \frac{\frac{1}{n}\left[\sum (N_h - 1) S_h^{\ 2} + \sum_h N_h (\bar{Y}_h - \bar{Y})^2 \right]}{N-1}$$

$$= \frac{\frac{1}{100}\left[686609.27 + 556547.18 \right]}{2009}$$

$$= 6.1879$$

Relative precision for proprotional allocation.

$$= \frac{\frac{1}{V_{prop}}}{\frac{1}{V_{ran}}} \times 100 = \frac{0.2913}{0.1616} \times 100$$

$$= 180\%$$

Relative precision for neyman allocation

$$= \frac{\frac{1}{V_{ney}}}{\frac{1}{V_{ran}}} \times 100 = \frac{6.1879}{2.8962} \times 100$$

$$= 213.655$$

$$= 214\%$$

$$Var(\underline{Y}_{st})_{prop} = \left(\frac{1}{n} - \frac{1}{N} \right) \sum P_i S_i^{\ 2}$$

$$= \left(\frac{1}{100} - \frac{1}{2010} \right) (689990.39)$$

$$= 6556.6250$$

$$Var(\overline{Y}_{st})_{opt} = \frac{1}{n}\left(\sum (N_h S_h)^2\right) - \frac{1}{N}(2N_h S_h^2)$$

$$= \frac{1}{100}(179795375.2) - \frac{1}{2010}(689990.39)$$

$$= 1797610.473$$

Proportional Allocation :

$$n_h = \left(\frac{n}{N}\right) N_h$$

$$n_1 = \frac{100}{2010} \times 394 = 19.6019$$

$$n_2 = \frac{100}{2010} \times 461 = 22.935$$

$$n_3 = \frac{100}{2010} \times 391 = 19.4527$$

$$n_4 = \frac{100}{2010} \times 334 = 16.6169$$

$$n_5 = \frac{100}{2010} \times 169 = 8.4079$$

$$n_6 = \frac{100}{2010} \times 113 = 5.6219$$

$$n_7 = \frac{100}{2010} \times 147 = 7.3631$$

Optimum Allocation :

$$n_h = \frac{n}{\sum N_h - S_h} \times N_h S_h$$

$$n_1 = \frac{100 \times 3270.20}{34206.9} = 9.5600$$

$$n_2 = \frac{100 \times 6131.3}{34206.9} = 17.9242$$

$$n_3 = \frac{100 \times 5904.1}{34206.9} = 17.2599$$

$$n_4 = \frac{100 \times 6613.2}{34206.9} = 19.3329$$

$$n_5 = \frac{100 \times 4140.5}{34206.9} = 12.1042$$

$$n_6 = \frac{100 \times 2938}{34206.9} = 8.5889$$

$$n_7 = \frac{100 \times 5209.6}{34206.9} = 15.2296$$

Conclusion :

Stratum	Proportional Allocation	Neyman Allocation
1	20	10
2	23	18
3	19	17
4	17	19
5	8	12
6	6	9
7	7	15
	100	100

$V_{prop} = 3.4327$

$V_{ney} = 2.8962$

$V_{ran} = 6.1879$

Relative precision for neyman allocation = 214%

Relative precision for proportional allocation = 180%

PRACTICAL - 2
Gain in Precision due to Stratification

The following date are desired from a stratified sample of tire dealers taken in March 2014. The dealers were assigned to strata according to the number of new tires held at a previous census. The sample mean $\bar{y}_h$ are the mean number of new per dealers.

a) Estimate the gain in precision due to stratification.

b) Compare the result with the gain that would have been attained from proportional allocation.

Stratum Boundries	N_h	$\bar{y}_h$	S_h^2 n_h	n_h
01-09	19850	4.1	34.8	3000
10-19	3250	13.0	92.2	600
20-29	1007	25.0	174.2	340
30-39	606	38.2	320.4	230
	24713			4170

Aim :

To estimate the gain in precision due to stratification also compare this result with the gain that would have been attained from proportional allocation.

Procedure :

$$V(\bar{y}_{st}) = \frac{\sum W_h^2 S_h^2}{n_h} - \frac{\sum W_h^2 S_h^2}{N_h}$$

Also,

$$V(\bar{y}_{st})_{prop} = \frac{N-n}{Nn} \sum_h W_h S_h^2$$

Now,

$$V(\bar{Y}) = \frac{N-n}{n(N-1)} \left[\sum_h W_h . S_h^2 + \sum_h W_h . \bar{Y}_h^2 - \left(\sum W_h . \bar{Y}_h \right)^2 \right]$$

Gain in precision due to stratification :

$$= \frac{1}{V(\bar{Y}_{st})} - \frac{1}{V(\bar{Y})}$$

Gain in precision due to proportional allocation.

$$= \frac{1}{V_{prop}(\bar{Y}_{st})} - \frac{1}{V(\bar{Y})}$$

Farm Size	N_h	$\overline{Y}_{N_h}$	S_h	$S_h^{\,2}$	$N_h S_h$	$N_h S_h^{\,2}$	$(N_h S_h)^2$	$N_h.\overline{Y}_{N_h}$	$\left(\overline{Y}_{N_h} - \overline{Y}\right)^2$	$N_h\left(\overline{Y}_{N_h} - \overline{Y}\right)^2$	$(N_h - 1)S_h^{\,2}$	
01-09	19850	4.1	34.8	3000	0.8032	0.6451	27.9513	13.5017	10.8445	3.2931	0.0074	0.0011
10-19	3250	13.0	92.2	600	0.1315	0.0172	12.1243	22.2235	2.9223	1.7095	0.0026	0.0004
20-29	1007	25.0	174.2	340	0.0407	0.0016	7.0899	25.4375	1.0353	1.0715	0.0008	0.0002
30-39	606	38.2	320.4	230	0.0245	0.0006	7.8498	35.7513	0.8759	0.9359	0.0008	0.0003
	N = 24713			$\sum n_h =$ 4170		$\sum W_h^{\,2} =$ 0.6645	$\sum W_h S_h^{\,2} =$ 55.0153	$\sum W_h \overline{Y}_h^{\,2} =$ 96.914	$\sum \left(W_h \overline{Y}_h\right)^2 =$ 15.678	$\sum W_h \overline{Y}_h =$ 6.956	$\sum \frac{W_h^2 S_h^2}{n_h} =$ 0.0116	$\sum \frac{W_h^2 S_h^2}{N_h} =$ 0.002

$$V(Y_{st}) = \sum\left(\frac{W_h^{\,2} S_h^{\,2}}{n_h}\right) - \sum\left(\frac{W_h^{\,2} S_h^{\,2}}{N_h}\right)$$

$$= 0.0116 - 0.002$$

$$= 0.0096$$

Also,

$$V_p(\bar{Y}_{st}) = \frac{N-n}{Nn}\sum W_h S_h^2$$

$$= \frac{24713-4170}{(24713)(4170)}(55.0153)$$

$$V_{prop}(\bar{Y}_{st}) = 0.01096$$

$$V(\bar{Y}) = \frac{N-n}{n(n-1)}\left[\sum_h W_h S_h^2 + \sum_h W_h \bar{Y}_h^2 - (\sum W_h \bar{Y}_h)^2\right]$$

$$= \frac{24713-4170}{4170(24713-1)}[55.0153+96.914-15.678]$$

$$V(\bar{Y}) = 0.0271$$

Gain in precisioni due to stratification.

$$= \frac{1}{V(\bar{Y}_{st})} - \frac{1}{V(\bar{Y})}$$

$$= \frac{1}{0.0096} - \frac{1}{0.0271}$$

$$= 67.2663$$

Gain in precision due to proportional allocation

$$= \frac{1}{V_{prop}(\bar{Y}_{st})} - \frac{1}{V(\bar{Y})}$$

$$= \frac{1}{0.01096} - \frac{1}{0.0271}$$

$$= 91.2408 - 36.9003$$

$$= 54.3405$$

Conclusion :

$$V(\bar{Y}_{st}) = 0.0096$$

$$V(\bar{Y}) = 0.0271$$

$$V(\bar{Y}_{st})_{prop} = 0.01096$$

Gain in precision due to stratification

= 67.2663

Gain in precision due to proportional allocation

= 54.3405

PRACTICAL - 3
P.P.S Sampling

Sample survey was conducted to study the yield of paddy in AP. Sample of 20 forms a total of 100 was taken with probability proportional to the under area paddy crop replacement method, the total area under paddy crop (x) 484.5 hectors. the area under crop (x) and yiedl (y) were noted in hectarews and quintals per the area under crop (x) and yiedl into (y) were noted in hectares and quintals per hectares and quintals per hectare respectively, the sample selected by the cumulative total method was

Area under Crop (x_i)	5.2	5.9	3.9	4.2	4.7	4.8	4.9	6.8	4.7	5.7
Yield of Crop (y_i)	28	29	30	22	24	25	28	37	26	32
Area under Crop (x_i)	5.2	4.9	4.0	1.3	7.4	7.4	4.8	6.2	6.2	
Yield of Crop (y_i)	38	31	16	06	61	61	29	47	47	

Aim :

1. To estimate the average yield per farm using PPS Sampling with replacement.
2. To estimate the gain in precision due to PPS Sampling over simple random sampling with replacement.

Procedure :

$$\hat{\bar{Y}}_{PPS} = \frac{1}{nN}\sum_{i=1}^{n}\left(\frac{Y_i}{Pi}\right)$$

$$\hat{Y}_{PPS} = N.\hat{\bar{Y}}_{PPS}$$

$$V\left(\hat{\bar{Y}}_{PPS}\right) = \frac{1}{n(n-1)N^2}\left[\sum_{i=1}\left(\frac{Y_i}{P_i}\right)^2 - n.\hat{Y}^2_{PPS}\right]$$

$$V\left(\hat{Y}_{PPS}\right) = N^2.V(\hat{\bar{Y}}_{PPS})$$

$$V_{PPS}(\hat{Y}_{SR}) = \frac{1}{n^2}\left[N.\sum_{i=1}^{n}\frac{Y_i^2}{P_i} - n.\hat{Y}^2_{PPS}\right] + \frac{1}{n}V(\hat{Y}_{PPS})$$

$$V_{PPS}(\hat{\bar{Y}}_{SR}) = \frac{1}{N^2}V_{PPS}(\hat{Y}_{SR})$$

Gain in precision

$$= \left[\frac{1}{V(\hat{Y}_{PPS})} - \frac{1}{V_{PPS}(\hat{Y}_{SR})}\right] \times 100$$

Calculations :

X_i	Y_i	$P_i = \frac{X_i}{X}$	$\frac{Y_i}{P_i}$	$\left(\frac{Y_i}{P_i}\right)^2$	$\frac{Y_i^2}{P_i}$
5.2	28	0.0107	2616.8224	6847759.473	73271.0280
5.9	29	0.0121	2396.6942	5744143.088	69504.1322
3.9	30	0.0080	3750	14062500	112500
4.2	22	0.0086	2558.1395	6544077.701	56279.0697
4.7	24	0.0097	2474.2268	6121798.258	59381.4433
4.8	25	0.0099	2525.2525	6376900.189	63131.3131
4.9	28	0.0101	2772.2772	7685520.874	77623.7623
6.8	37	0.0140	2642.8571	6984693.651	97785.7142
4.7	26	0.0097	2680.4123	7184610.098	69690.7216
5.7	32	0.0117	2735.0427	7480458.571	87521.3675
5.2	25	0.0107	2336.4485	5458991.593	58411.2149
5.2	38	0.0107	3551.4018	12612454.75	134953.271
4.9	31	0.0101	3069.3069	9420644.846	95148.5148
4.0	16	0.0082	1951.2195	3807234.123	31219.5122
1.3	06	0.0026	2307.6923	5325443.751	13846.1538
7.4	61	0.0152	4013.1578	16105435.53	244802.6316
7.4	61	0.0152	4013.1578	16105435.53	244802.6316
4.8	29	0.0099	2929.2929	8580756.894	84949.4949
6.2	47	0.0127	3700.7874	13695827.38	173937.0079
6.2	47	0.0127	3700.7874	13695827.38	173937.0079
			58724.977	179840513.7	2022695.993

We know that,

$$\hat{\bar{Y}}_{PPS} = \frac{1}{nN}\sum_{i=1}^{n}\left(\frac{Y_i}{P_i}\right)$$

$$= \frac{1}{20\times 100}(58724.977)$$

$$\hat{\bar{Y}}_{PPS} = 29.3624$$

$$\hat{Y}_{PPS} = N.\hat{\bar{Y}}_{PPS} = 100\times(29.3624)$$

$$= 2936.24$$

$$V\left(\hat{\bar{Y}}_{PPS}\right) = \frac{1}{n(n-1)N^2}\left[\sum_{i=1}\left(\frac{Y_i}{P_i}\right)^2 - n.\hat{Y}_{PPS}^2\right]$$

$$= \frac{1}{20(19)(100)^2}\left[179840513.7 - 20(2936.24)^2\right]$$

$$V\left(\hat{\bar{Y}}_{PPS}\right) = 1.9501$$

$$V\left(\hat{Y}_{PPS}\right) = N^2.V\left(\hat{\bar{Y}}_{PPS}\right)$$

$$= (100)^2(1.9501)$$

$$V\left(\hat{Y}_{PPS}\right) = 19501$$

Also,

$$V_{PPS}\left(\hat{Y}_{SR}\right) = \frac{1}{n^2}\left[N\sum_{i=1}^{n}\left(\frac{Y_i}{P_i}\right)^2 - n.\hat{Y}^2_{PPS}\right] + \frac{1}{n}V\left(\hat{Y}_{PPS}\right)$$

$$= \frac{1}{(20)^2}\left[100(2022695.993) - 20(2936.24)^2\right] + \frac{1}{20}(1.9501)$$

$$= \frac{1}{400}\left[29839492.55\right] + 0.0975$$

$$V_{PPS}(\hat{Y}_{SR}) = 74598.8288$$

$$V_{PPS}\left(\hat{\bar{Y}}_{SR}\right) = \frac{1}{N^2}V_{PPS}\left(\hat{Y}_{SR}\right)$$

$$= \frac{1}{(100)^2}(74598.8288)$$

$$V_{PPS}\left(\hat{\bar{Y}}_{SR}\right) = 7.45988$$

Now,

Gain in precision = $\left[\frac{1}{V\left(\hat{\bar{Y}}_{PPS}\right)} - \frac{1}{V_{PPS}\left(\hat{\bar{Y}}_{SR}\right)}\right] \times 100$

$$= \left[\frac{1}{1.9501} - \frac{1}{7.4598}\right] \times 100$$

$$= 37.8742$$

Conclusion :

$$\hat{\bar{Y}}_{PPS} = 29.3624$$

$$\hat{\bar{Y}}_{PPS} = 2936.24$$

$$V\left(\hat{\bar{Y}}_{PPS}\right) = 1.9501$$

$$V\left(\hat{Y}_{PPS}\right) = 19501$$

$$V_{PPS}\left(\hat{Y}_{SR}\right) = 74598.8288$$

$$V_{PPS}\left(\hat{\bar{Y}}_{SR}\right) = 7.4598$$

$\therefore$ Gain in precision = 37.8742

PRACTICAL - 4
Ratio Method of Estimation

A sample of 34 villages was selected from a population of 170 villages in a region.. The following table gives the data of cultivated area under wheat in 2003(x) and in 2004(y) for these sample villages.

S.No.	1	2	3	4	5	6	7	8	9	10	11
X	70	163	320	440	250	125	558	254	101	359	109
Y	50	149	284	381	278	111	634	278	112	355	99

12	13	14	15	16	17	18	19	20	21	22	23
481	125	5	427	78	75	45	546	238	91	247	134
498	111	6	339	80	105	27	515	241	85	221	133

24	25	26	27	28	29	30	31	32	33	34
131	129	190	363	235	73	62	71	137	196	255
144	103	175	335	219	62	79	60	100	141	263

i) Estimate the area under wheat in 2004 by the method of ratio estimation using information on wheat area as x = 21288 acres for 2023.

ii) Determine the efficiency the ratio estimate as compared to the usual simple random sample estimate.

Aim :

i) To estimate the area under wheat in 2004 by the method of ratio estimation using information on wheat areas as x = 21288 acres for 2003 and

ii) To determine the efficiency the ratio estimate as compared to the usual simple random simple estimate.

Procedure :

$$\hat{Y}_R = \frac{\overline{Y}}{\overline{X}}.X$$

Where, $\overline{Y}$ and $\overline{X}$ are sample mean.

$$V\left(\hat{Y}_R\right) = \frac{N(N-n)}{n(n-1)}\left[\sum_{i=1}^{n} Y_i^2 + \hat{R}^2 \sum_{i=1}^{n} X_i^2 - 2.\hat{R}\sum_{i=1}^{n} X_i Y_i\right]$$

Where, $\hat{R} = \frac{\overline{Y}}{\overline{X}}$

$$V\left(\hat{Y}\right) = N^2\left(\frac{N-n}{N}\right).\frac{SY^2}{n}$$

Where, $S_Y^2 = \sum_{i=1}^{n} \frac{\left(Y_i - \overline{Y}\right)^2}{n-1}$

Relative efficiency = $\frac{V\left(\hat{Y}\right)}{V\left(\hat{Y}_R\right)} \times 100$

Calculations :

X_i	Y_i	X_i^2	Y_i^2	X_iY_i	$(Y_i - \bar{Y})^2$
70	50	4900	2500	3500	22262.3707
163	149	26569	22201	24287	2520.6223
320	284	102400	80656	90880	7190.0563
440	381	193600	145161	167640	33049.1311
250	278	62500	77284	69500	6208.5259
125	111	15625	12321	13875	7780.2631
558	634	311364	401956	353772	189045.9964
254	278	64516	77284	70612	6208.5259
101	112	10201	12544	11312	7604.8515
359	355	128881	126025	127445	24271.8327
109	99	11881	9801	10791	10041.2023
481	498	231361	248004	239538	89277.9739
125	111	15625	12321	13875	7780.2631
5	6	25	36	30	37328.4811
427	339	182329	114921	144753	19542.4183
78	80	6084	6400	6240	14210.0227
75	105	5625	11025	7875	8874.7327
45	27	2025	729	1215	29654.8375
564	515	318096	265225	290460	99725.9767
238	241	56644	58081	57358	1746.7551
92	85	8464	7225	7820	13042.9647
247	221	61009	48841	54587	474.9871
134	133	17956	17689	17822	4383.2079
131	144	17161	20736	18864	3047.6803
129	103	16641	10609	13287	9255.5559
190	175	36100	30625	33250	585.9207
363	335	131769	112225	121605	18440.0647
235	219	55225	47961	51465	391.8103
73	62	5329	3844	4526	18825.4315
62	79	3844	6241	4898	14449.4343
71	60	5041	3600	4260	19378.2547

X_i	Y_i	X_i^2	Y_i^2	X_iY_i	$(Y_i - \bar{Y})^2$
137	100	18769	10000	13700	9841.7907
196	141	38416	19881	27636	3387.9151
255	263	65025	69169	67065	4069.6999
$\sum X_i =$ 7102	$\sum Y_i =$ 6773	$\sum X_i^2 =$ 2231000	$\sum Y_i^2 =$ 2093121	$\sum X_iY_i =$ 2145743	$\sum (Y_i - \bar{Y})^2 =$ 743900.5571

$$\Rightarrow \bar{Y} = \frac{1}{n}\sum Y_i = \frac{6773}{34} = 199.2058$$

$$\Rightarrow \bar{X} = \frac{1}{n}\sum X_i = \frac{7025}{34} = 208.8823$$

$$\hat{R} = \frac{\bar{Y}}{\bar{X}} = \frac{199.2058}{208.8823} = 0.9536$$

$$\hat{Y}_R = \frac{\bar{Y}}{\bar{X}}.X = (0.9536) \times 21288$$

$$= 20300.2368$$

Now,

$$V\left(\hat{Y}_R\right) = \frac{N(N-n)}{n(n-1)}\left[\sum_{i=1}^{n} Y_i^2 + \hat{R}^2 \sum_{i=1}^{n} X_i^2 - 2\hat{R}\sum_{i=1}^{n} X_iY_i\right]$$

$$= \frac{170(170-34)}{34(34-1)}\left[2093121 + (0.9536)^2(2231000) - 2(0.9536)(2145743)\right]$$

$$= (20.6060)\left[2093121 + 2028766.454 - 4092361.05\right]$$

$$V(\hat{Y}_R) = 608421.0808$$

$$\Rightarrow V(\hat{Y}) = N^2\left(\frac{N-n}{N}\right).\frac{SY^2}{n}$$

$$S_Y^2 = \sum_{i=1}^{n}\frac{\left(Y_i - \bar{Y}\right)^2}{n-1} = \frac{743900.5571}{34-1}$$

$$S_Y^2 = 22542.4411$$

Now,

$$V(\hat{Y}) = (170)^2\left(\frac{170-34}{170}\right)\left(\frac{22542.4411}{34}\right)$$

$$= (28900)(0.8)\ (663.0129)$$

$$V(Y) = 15328858.25$$

Relative Efficiency

$$= \frac{V(\hat{Y})}{V(\hat{Y}_R)} \times 100$$

$$= \frac{(15328858.25)}{(608421.0808)} \times 100$$

$$= 2519.4489$$

Conclusion :

$\hat{Y}_R = 20300.2368$

$V(\hat{Y}_R) = 608421.0808$

$V(\hat{Y}) = 15328858.25$

Relative Efficiency = 2519.4489

PRACTICAL - 5
Regression Method of Estimation

An experienced farmer makes an estimate of the weight of apples on each tree. In an Orchard of N=200 trees. He finds an total weight of x = 11,600 lbs. The apples are picked and weighted on a SRS of 10 trees with the following results :

	Tree Numbers									
	1	2	3	4	5	6	7	8	9	10
Actual Weight (Y_i)	61	42	50	58	67	45	39	57	71	53
Estimated Weight (X_i)	59	47	52	60	67	48	44	58	76	58

Compute the regression estimate for the total acutal weight 'Y' and find its standard error.

Aim :

To compute the regression estimate for the total actual weight 'Y' and find its standard error.

Procedure :

Regression estimate of the population mean is

$$\overline{Y}_{lr} = \overline{Y}_{lr} = \overline{Y} + b(\overline{X} - \overline{X})$$

Where,

$$b = \frac{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)\left(X_i - \overline{X}\right)}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$$

$$= \frac{\sum_{i=1}^{n} X_i Y_i - n.\overline{X}\,\overline{Y}}{\sum_{i=1}^{n} X_i^2 - n.\overline{X}^2}$$

Now,

$$\overline{Y}_{lr} = N.\hat{\overline{Y}}_{lr} = N.\overline{Y}_{lr} \longrightarrow$$

$$V\left(\overline{Y}_{lr}\right) = \frac{1-f}{n(n-2)} \sum_{i=1}^{n} \left[\left(Y_i - \overline{Y}\right) - b(X_i - \overline{X})\right]^2$$

$$= \frac{N-n}{Nn(n-2)}\left[\sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)^2 - \frac{\left[\sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)\left(X_i-\bar{X}\right)\right]^2}{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}\right]$$

then,

$$V\left(\hat{Y}_{lr}\right)=N^2.V\left(\hat{\bar{Y}}_{lr}\right)$$

Standard error of $\hat{Y}_{lr} = \sqrt{V\left(\hat{Y}_{lr}\right)} \longrightarrow$

Calculations :

S.No.	X_i	Y_i	X_i^2	Y_i^2	X_iY_i	$(Y_i-\bar{Y})^2$	$(X_i-\bar{X})^2$	$(Y_i-\bar{Y})(X_i-\bar{X})$
1	59	61	3481	3721	3599	44.89	4.41	14.07
2	47	42	2209	1764	1974	151.29	98.01	121.77
3	52	50	2704	2500	2600	18.49	24.01	21.07
4	60	58	3600	3364	3480	13.69	9.61	11.47
5	67	67	4489	4489	4489	161.29	102.01	128.27
6	48	45	2304	2025	2160	86.49	79.21	82.77
7	44	39	1936	1521	1716	234.09	166.41	197.37
8	58	57	3364	3249	3306	7.29	1.21	2.97
9	76	71	5776	5041	5796	278.89	364.81	318.97
10	58	53	3364	2809	3074	1.69	1.21	-1.43
	$\sum X_i=$ 569	$\sum Y_i=$ 543	$\sum X_i^2=$ 33227	$\sum Y_i^2=$ 30483	$\sum X_iY_i=$ 31794	$\sum(Y_i-\bar{Y})^2=$ 998.1	$\sum(X_i-\bar{X})^2=$ 850.9	$\sum(Y_i-\bar{Y})(X_i-\bar{X})=$ 897.3

Now,

Given that

$$\bar{X}=\frac{1}{n}\sum X_i=\frac{569}{10}=56.9$$

$$\bar{Y}=\frac{1}{n}\sum Y_i=\frac{543}{10}=54.3$$

Then,

$$\overline{Y}_{lr} = \overline{Y} + b(\overline{X} - \overline{X})$$

Where

$$b = \frac{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)\left(X_i - \overline{X}\right)}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$$

$$= \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{X}\overline{Y}}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}$$

$$= \frac{31794 - 10(56.9)(54.3)}{33227 - 10(56.9)^2}$$

$$= \frac{897.3}{850.9}$$

$$\overline{Y}_{lr} = \overline{Y} + b(\overline{X} - \overline{X})$$

$$= 54.3 + (1.0545)(58 - 56.9)$$

$$\overline{Y}_{lr} = 55.4599$$

Also,

$$\overline{Y}_{lr} = N.\overline{Y}_{lr}$$

$$= 200(55.4599)$$

$$\overline{Y}_{lr} = 11091.98$$

$$\Rightarrow V\left(Y_{lr}\right) = \frac{N-n}{Nn(n-2)}\left[\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2 - \frac{\left[\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)\left(X_i - \overline{X}\right)\right]^2}{\sum_{i=1}^{n}\left(X_i - \hat{X}\right)^2}\right]$$

$$= \frac{200-10}{200(10)(8)}\left[998.1 - \frac{805147.29}{850.9}\right]$$

$$= \frac{190}{16000}\left[998.1 - 946.2302\right]$$

$$= (51.8698)(0.0118)$$

$$V\left(\overline{Y}_{lr}\right) = 0.6120$$

$$V\left(\overline{Y}_{lr}\right) = N^2.V\left(\overline{Y}_{lr}\right)$$

$$= (200)^2)(0.6110)$$

$$V\left(\overline{Y}_{lr}\right) = 24480$$

Standard Error of :

$$\overline{Y}_{lr} = \sqrt{V\left(\hat{\overline{Y}}_{lr}\right)}$$

$$= \sqrt{24480}$$

$$\overline{Y}_{lr} = 156.4608$$

Conclusion :

Here, the values are

b = 1.0545

$$\overline{Y}_{lr} = 55.4599$$

$$\hat{\overline{Y}}_{lr} = 11091.98$$

$$V\left(\hat{\overline{Y}}_{lr}\right) = 0.6120$$

$$V\left(\hat{Y}_{lr}\right) = 24480$$

$-$

PRACTICAL - 6
Gain Efficiency in Stratified Random Sampling

Below is a structured response for the stratified random sampling example (estimating average student test scores) formatted with Aim, Procedure, Calculation, and Inference. The example follows the same scenario: a school with 1,000 students across 9th (400), 10th (350), and 11th (250) grades, with known standard deviations (10, 15, 20) and a sample size of 100.

Aim :

To estimate the average test score of 1,000 students in a school using stratified random sampling, ensuring efficiency by minimizing the variance of the estimate compared to simple random sampling.

Procedure:

1. **Identify Strata :**

- Divide the population into three strata based on grade : 9th (400 students), 10th (350 students), 11th (250 students).
- Note population sizes $(N_1 = 400,\ N_2 = 350,\ N_3 = 250)$ and standard deviation $(\sigma_1 = 10, \sigma_2 = 15, \sigma_3 = 20)$.

2. **Determine Sample Size :**

- Total Sample size : $\eta = 100$.

3. **Allocate Sample Using Neyman Allocation :**

- Allocate samples to each stratum proportionally to $N_h.\sigma_h$ to minimize variance.
- Formula: $n_h = n\dfrac{N_h - \sigma_h}{\sum(N_h - \sigma_h)}$

4. **Selected Samples :**

- Randomly select the calculated number of students from each stratum using a random number generator or sampling software.

5. **Collect Date :**

- Record test scores for the sampled students in each stratum.

6. **Estimate Population Mean :**

- Calculate the stratum means and compute the weighted average using population proportions $\mu - \sum\left(\dfrac{N_h}{N}.y_h\right)$.

7. **Assess Effieciency :**

- Calculate the variance of the stratified estimate and compare it to the variance of a simple random sampling estimate to confirm efficiency.

Calculation :

Step 1: Allocate Sample Sizes (Neyman Allocation)

- Population Sizes : $N_1 = 400,\ N_2 = 350,\ N_3 = 250$
- Standard Deviations : $\sigma_1 = 10, \sigma_2 = 15, \sigma_3 = 20$
- Compute $N_h.\sigma_h$.

- 9th grade : 400 . 10 = 400
- 10th grade : 350. 15 = 5250
- 11th grade : 250 . 20 = 5000
- Sum : 4,000 + 5,250 + 5,000 = 14250.
- Sample Sizes :

9th grade : $n_1 = 100.\frac{4000}{14250} = 28.07 \approx 28$

10th grade : $n_2 = 100.\frac{5250}{14250} = 36.84 \approx 36$

11th grade : $n_3 = 100.\frac{5000}{14250} = 35.09 \approx 35$

- Total : 28 + 37 + 35 = 100.

Step 2 : Collect Sample Data (Hypothetical)

- Assume random sampling yields.
  - $\text{9th grade}(28 \text{ students}):\ \text{Mean Score } \overline{\text{y}}_1 = 85$
  - $\text{10th grade}(37 \text{ students}):\ \text{Mean Score } \overline{\text{y}}_2 = 80$
  - $\text{11th grade }(35 \text{ students}):\ \text{Mean Score } \overline{\text{y}}_3 = 78$

Step 3 : Estimate Population Mean

- Total population : $N = 1000$
- Weights :
  - 9th grade: $\frac{\text{N}_1}{\text{N}} = \frac{400}{1000} = 0.4$
  - 10th grade: $\frac{\text{N}_2}{\text{N}} = \frac{350}{1000} = 0.35$
  - 11th grade: $\frac{\text{N}_3}{\text{N}} = \frac{250}{1000} = 0.25$
- Mean estimate :

$\bar{\mu} - (0.4 * .85) + (0.35 * .80) + (0.25 * .78) - 34 + 28 + 19.5 = 81.5$

Step 4 : Calculate Variance (Stratified Sampling)

- Variance formula :

$$Var(\bar{\mu}_{strat}) = \sum \left(\frac{N_h}{N}\right)^2 . \frac{\sigma_h^2}{n_h} . \frac{N_h - n_h}{N_h}$$

- 9th grade :

$$\left(\frac{400}{1000}\right)^2 . \frac{10^2}{28} . \frac{400 - 28}{400} - 0.16.\frac{100}{28}.0.93 \approx 0.531$$

- 10th grade :

$$\left(\frac{350}{1000}\right)^2.\frac{15^2}{37}.\frac{350-37}{350}-0.1225.\frac{225}{37}.0.8943\approx 0.667$$

- 11th grade :

$$\left(\frac{250}{1000}\right)^2.\frac{20^2}{35}.\frac{250-35}{250}-0.0625.\frac{400}{35}.0.86\approx 0.614$$

- Total variance :

$$Var\left(\bar{\mu}_{strat}\right)=0.531+0.667+0.614\approx 1.812$$

Step 5 : Calculate Variance (Simple Random Sampling)

- Population variance:

$$\sigma^2=\sum\left(\frac{N_h}{N}.\sigma_h^2\right)-\left(0.4*.10^2\right)+\left(0.35*.15^2\right)+\left(0.25*.20^2\right)=40+78.75+100=218.75$$

- SRS variance:

$$Var\left(\bar{\mu}_{SRS}\right)=\frac{\sigma^2}{n}.\frac{N-n}{N}-\frac{218.75}{100}.\frac{1000-100}{1000}-2.1875*0.9\approx 1.969$$

- Standard error :

$$\sqrt{1.969}\approx 1.403$$

Step 6 : Efficiency Comparison

- Variance reduction: $\dfrac{1.969-1812}{1.969}\approx 0.0797$ or 7.97%

- Stratified sampling reduces variance by 8%.

Conclusion:

Stratified random sampling, applied to estimate the average test score of 1,000 students across three grades, yields an efficient and precise estimate of 81.5 with a variances of 1.812, approximately 8% lower than the variance of simple random sampling (1.969). By using Neyman allocation to assign samples (28, 37 and 35 students to 9th, 10th, and 11th grades respectively), the method optimized precision by accounting for stratum size and variability. This ensures proportional representation and minimizes sampling error, making stratified random sampling superior for heterogeneous populations. There approach is practical and effective, provided stratum characteristics are accurately known, offering a cost efficient way to achieve reliable estimates with smaller sample sizes.

PRACTICAL – 7

Precision of Systematic Sampling and Stratified Sampling

The data below are a small artificial population which exhibits a fairly study rising trend. Each column represents a systematic sample and the row that strata compare the precision of systematic sampling and stratified sampling. The data for 10 systematic samples with n=4, k=10 and nk=N=40.

Strata	Systematic									
	1	2	3	4	5	6	7	8	9	10
I	0	1	1	2	5	4	7	7	8	6
II	6	8	9	10	13	12	15	16	16	17
III	18	19	20	20	24	23	25	28	29	28
IV	26	30	31	31	33	32	35	37	38	38

Aim :

To compare ethe precision of systematic sampling and stratified sampling.

Formula and Procedure :

$$V(\bar{Y}) = \frac{1}{n}\sum_{i=1}^{K}\left(\bar{Y}_i - \bar{Y}\right)^2$$

$$= \frac{1}{K}\left(\sum \bar{Y}_i^2 - K\bar{Y}\right)^2$$

$$= \frac{1}{n^2 K}\left(\sum_{i=1}^{K}\left(n\bar{Y}_i\right)^2 - n^2 K\bar{Y}^2\right)$$

$$= \frac{K-1}{nK} S^2_{wst}$$

Where,

$$S^2_{wst} = \frac{1}{n}(K-1)\sum_{i=1}^{K}\sum_{j=1}^{K}\left(Y_{ij} - \bar{Y}_{.j}\right)$$ is within (stratun)

Mean sum of squares

$$V(\underline{Y})_{n\ R} = \left(\frac{1}{n} - \frac{1}{N}\right)S^2$$

$$S^2 = 1(N-1)\sum\sum\left(Y_{ij} - \bar{Y}_{..}\right)^2$$ is total mean sum of squares

Correlation factor (F) = $\frac{G^2}{N}$

Total sum of squares (rss) = $\sum_{i=1}^{n}\sum_{j=1}^{n} Y_{ij}^{2} - CF$

Between the strata and sum of squares is

$$\sum_{i=1}^{n} \frac{T_j^{2}}{K} - CF$$

Within strata sum of squares - Total sum of squares;

$$S_{wst}^{2} = \frac{within\ strata\ SS}{n(K-1)}$$

$$S^{2} = \frac{Total\ sum\ of\ squares}{N-1}$$

ANDVA Table :

Source of Variation	df	Sum of Squares	Mean Sum of Squares
Between strata	n-1	Between Strata S.S	-
Witin Strata	n(K-1)	Within Strata SS	S^2_{wst}
Total	N-1	T.SS	S^2

Calculations :

We have

$$V(\bar{Y}_{st}) = \frac{1}{n^2K}\left[\sum_{i=1}^{K}\left(n\bar{Y}_i\right)^2 - n^2K\bar{Y}^2\right]$$

Strata	Systematic Sampling Number										Strata
	1	2	3	4	5	6	7	8	9	10	Total (T_j)
I	0	1	1	2	5	4	7	7	8	6	41
II	6	8	9	10	13	12	15	16	16	17	122
III	18	19	20	20	24	23	25	28	29	27	233
IV	26	30	31	31	33	32	35	37	38	38	331
Total $n(\bar{Y}_i)$	50	58	61	63	75	71	82	88	91	88	727

$$CF = \frac{G^2}{N} = \frac{(727)^2}{40} = 13213.225$$

Total sum of squares = $\sum_{i=1}^{K}\sum_{j=1}^{K} Y_{ij} - CF$

$$= (o^2 + 1^a + 2^2 + ... + 38^2) - (13213.225)$$

$$= 18527 - 13213.225$$

$$= 5313.775$$

Between Strata Sum of Squares = $\sum_{i=1}^{n} \frac{T_i^2}{K} - CF$

$$= \frac{1}{10}\left[(41)^2 + (122)^2 + (233)^2 + (331)^2\right] - 13213.225$$

$$= 4828.275$$

Witin in Strata Sum of Squares = TSS - Between Strata SS

= 5313.775 - 4828.275

= 485.5

$$S^2{}_{wst} = \frac{within\ Strata\ SS}{n(K-1)}$$

$$= \frac{485.5}{4(9)}$$

$$= 13.4861$$

$$S^2 = \frac{Total\ Sum\ of\ Squares}{N-1}$$

$$= \frac{5313.775}{39}$$

$$S^2 = 136.2506$$

Source of Variation	df	Sum of Squares	Mean Sum of Squares
Between strata	(4-1) = 3	4828.275	-
Witin Strata	4(10-1) = 36	485.5	13.4861
Total	39	5313.775	136.2506

$$V(\overline{Y}_{sys}) = \frac{1}{n^2K}\left[\sum_{i=1}^{K}\left(nY_i\right)^2 - \left(n^2 KY..\right)^2\right]$$

$$\overline{Y}^{-2}_{..} = \frac{1}{K}\sum_{i=1}^{K}\overline{Y}_i$$

$$= \frac{1}{10}\left[\frac{727}{4}\right]$$

$$\bar{Y}_{..}^{2} = 18.175$$

$$V(\bar{Y}_{sys}) = \frac{1}{16(10)}\left[54713 - 160(18.175)^2\right]$$

$$= \frac{1}{160}(1860.1)$$

$$= 11.6256$$

$$V(\bar{Y}_n)_R = \left(\frac{1}{n} - \frac{1}{N}\right)S^2 = \left(\frac{1}{4} - \frac{1}{40}\right)(136.2506)$$

$$= (0.225)(136.2506)$$

$$V(\bar{Y}_n)_R = 30.6563$$

$$V(\bar{Y}_{st}) = \frac{K-1}{nK} S^2_{wst}$$

$$= \frac{10-1}{4(10)}(13.4861)$$

$$= \frac{9}{40}(13.4861)$$

$$V(\bar{Y}_{st}) = 3.0343$$

Coclusion :

$$(\bar{Y}_{..}^{2}) = 18.175$$

$$V(\bar{Y}_{sys}) = 11.6256$$

$$V(\bar{Y}_n)_R = 30.6563$$

$$V(\bar{Y}_{st}) = 3.0343$$

It may be observed that

$$V(\bar{Y}_{st}) < V(\bar{Y}_{sys}) < V(\bar{Y}_n)_R$$

$\therefore$ Since, the data exists a fairly study rising trend.

PRACTICAL – 8

Unbiased Estimator of the Population Mean for Systematic Sampling

Strata	Systematic Sample Numbers				
	1	2	3	4	5
I	65	70	68	65	60
II	60	58	63	59	55
III	61	56	68	56	51
IV	63	60	67	62	58

In the above table each column represents a systematic sample and the rows are the strata :

i) Prove that the sample mean is unbiased estimator of population mean.

ii) Prove that $V(\bar{Y}_{sys}) = \left(\frac{N-1}{N}\right)S^2 - \left(\frac{n-1}{n}\right)S^2_{wsy}$

Aim :

To calculate the sample mean is an unbiased estimator of the population mean for systematic sampling and also obtain the

$$V(\bar{Y}_{sys}) = \left(\frac{N-1}{N}\right)S^2 - \left(\frac{n-1}{n}\right)S^2_{wsy}$$

Procedure :

We know that,

Sample mean $E(\bar{Y}_{i.}) = \frac{1}{K}\sum_{i=1}^{K}\bar{Y}_i$

Population mean $\bar{Y}_{..} = \frac{1}{N}\sum_{i=1}^{K}\sum_{j=1}^{n}Y_{ij}$

$$V(\bar{Y}_{sys}) = \left(\frac{N-1}{N}\right)S^2 - \left(\frac{n-1}{n}\right)S^2_{wsy}$$

Where,

$S^2 = Population\ mean\ square$

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{K}\sum_{j=1}^{n}\left(Y_{ij} - \bar{Y}_{..}\right)^2$$

$S^2_{wsy} = Sample\ mean\ square\ when\ sample\ systematic\ sample$

$$= \frac{1}{K(n-1)}\sum_{i=1}^{K}\sum_{j=1}^{n}\left(Y_{ij} - \bar{Y}_{i.}\right)^2$$

Calculation :

Here N=20; n=5 and K=4

Now we have to find

$$E\left(\bar{Y}_{i.}\right)=\bar{Y}_{..}$$

Take $E\left(\bar{Y}_{i.}\right)=\frac{1}{K}\sum \bar{Y}_{i.}$

$$=\frac{1}{4}(245)$$

$$E\left(\bar{Y}_{i.}\right)=61.25$$

$$\Rightarrow \bar{Y}_{..}=\frac{1}{N}\sum_{i=1}^{K}\sum_{j=1}^{n}Y_{ij}$$

$$=\frac{1}{20}\times(1225)$$

$$\bar{Y}_{..}=61.25$$

$$\therefore E\left(\bar{Y}_{i.}\right)=\bar{Y}_{..}$$

i) To prove

$$V(\bar{Y}_{sys})=\left(\frac{N-1}{N}\right)S^2-\left(\frac{n-1}{n}\right)S^2_{wsy}$$

$$S^2=\frac{1}{N-1}\sum_{i=1}^{K}\sum_{j=1}^{n}\left(Y_{ij}-\bar{Y}_{..}\right)^2$$

$$S^2=\frac{1}{20-1}[14.0625+76.5625+45.5625+14.0625+1.5625+1.5625+10.5625+3.0625+5.0625+39.0625+0.0625+27.5625+45.5625+27.5625+105.0625+3.0625+1.5625+33.0625+0.5625+10.5625]$$

$$=\frac{1}{19}[465.6875]$$

$$S^2=24.5098$$

$$S^2_{wsy}=\frac{1}{K(n-1)}\left[\sum_{i=1}^{K}\sum_{j=1}^{n}\left(Y_{ij}-\bar{Y}_{i.}\right)^2\right]$$

$$=\frac{1}{4(4)}[0.36+19.36+5.76+31.36+0.36+1+1+16+0+16+6.76+5.76+92.16+5.76+54.76+1+4+25+0+16]$$

$$=\frac{1}{16}[302.4]$$

$$S^2_{wsy} = 18.9$$

$$\therefore V(\overline{Y}_{sys}) = \left(\frac{N-1}{N}\right)S^2 - \left(\frac{n-1}{n}\right)S^2_{wsy}$$

$$= \left(\frac{19}{20}\right)(24.5098) - \left(\frac{4}{5}\right)(18.9)$$

$$= 23.2843 - 15.12$$

$$V(\overline{Y}_{sys}) = 8.1643$$

Conclusion :

i) Sample mean is an unbiased estimator of population mean;

i.e., $E(\overline{Y}_{i.}) = \overline{Y}_{..}$

ii) $V(\overline{Y}_{sys}) = 8.1643$

PRACTICAL – 9

Yield of Average and Standard Error

At an experimental station there were 100 fields with wheat. Each field was divided into 16 plots, of equal size (1/6th hectared) out of 100 tickets 10 were selected by SRSWOR for each selected field, 4 plots were chosen by random sampling WOR the yield in kg/plot are given below.

Selected Field	Plots			
	1	2	3	4
1	4.32	4.84	3.96	4.04
2	4.16	4.36	3.50	5.00
3	3.06	4.24	4.76	3.12
4	4.00	4.84	4.32	3.72
5	4.12	4.60	3.46	4.02
6	4.08	3.96	3.42	3.08
7	5.16	4.24	4.96	3.84
8	4.40	4.72	4.04	3.98
9	4.20	4.66	3.64	5.00
10	4.28	4.36	3.00	3.52

i) Estimate the wheat yield per hectare for the experiments stations along with its SE.

ii) How can you estimator obtained from a simple random sample of 40 plots be compared with estimate. Obtain above in (i) :

iii) Obtain Optimum ‘n’ and ‘m’ under constant function 100 = 4n + nm.

Aim :

i) To estimate that the wheat yield of average and SE.

ii) To estimate ‘n’ and m under the cast function.

Procedure :

$$\bar{Y} = \frac{1}{n}.\sum_{i=1}^{n} \bar{Y}_{i.}$$

To estimate $V(\bar{Y})$

$$V\left(\underline{Y}\right)=\left(\frac{1}{n}-\frac{1}{N}\right)S^2{}_b+\frac{1}{n}\left(\frac{1}{m}-\frac{1}{M}\right)S^2{}_W$$

$$S^2{}_b=\frac{1}{n-1}\sum_{i=1}^{n}\left(\overline{Y}_{i.}-\overline{Y}\right)^2$$

$$S^2{}_W=\frac{1}{n(m-1)}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(Y_{ij}-\overline{Y}_{i.}\right)^2$$

In SRS, the estimate of varianee is given by

$$V_{ran}\left(\underline{Y}\right)=\left(\frac{1}{nm}-\frac{1}{NM}\right)S^2$$

Cost function C = Gn + C_2nm

The Optimum value

$$m_{opt}=\left(\frac{C_1S_w^2}{C_2S_b^2-\frac{S_W^2}{m}}\right)^{1/2}$$

$$S^2=\frac{1}{NM-1}\left[M(N-1)S_b^2+\left\{N(m-1)-(M-m)\frac{N-1}{m}\right\}S_W^2\right]$$

$$V(\underline{Y})=\left(\frac{1}{n}-\frac{1}{N}\right)S_b^2+\frac{1}{n}\left(\frac{1}{m}-\frac{1}{M}\right).S_W^2$$

and S.E of $\overline{Y}=\sqrt{V(\overline{Y})}$

i) In sample random sampling the estimate of variance is given by :

$$V_{ran}(Y)=\left(\frac{1}{nm}-\frac{1}{NM}\right)S^2$$

The estimate of S_i^2 using a two - stage sampling ncan be written as;

$$S^2=\frac{1}{NM-1}\left[M(N-1)S_b^2+\left\{N(m-1)-(M-m)\frac{N-1}{m}\right\}S^2{}_w\right]$$

Calculations :

S.No.	X_i	Y_i	X_i^2	Y_i^2	X_iY_i	$(Y_i - \bar{Y})^2$
1	17.16	4.29	0.0273	74.0912	18.4041	0.4748
2	17.02	4.255	0.0170	73.5652	18.1050	1.1452
3	15.18	3.795	0.1085	59.7332	14.4020	2.1252
4	16.88	4.22	0.0091	71.9264	17.8084	0.6928
5	16.2	4.05	0.0055	66.2664	16.4025	0.6564
6	14.54	3.635	0.2396	53.5108	13.2132	0.658
7	18.20	4.55	0.1810	83.9504	20.7025	1.1404
8	17.14	4.285	0.0257	73.8004	18.3612	0.3556
9	17.50	4.375	0.0627	77.6052	19.1406	1.0428
10	15.16	3.79	0.1118	58.7184	14.3641	1.262
		$\sum \bar{Y}_{i.} =$ 41.245	$\sum(\bar{Y}_i - \bar{Y})^2 =$ 0.7882	$\sum(\bar{Y}_i - \bar{Y})^2 =$ 693.1676	$\sum \bar{Y}_i^2 =$ 170.9036	$\sum\left(\sum_{j=1}^{n} Y_{ij}^2 - m\bar{Y}_{i.}^2\right) =$ 9.5532

$$\bar{Y} = \frac{1}{n}\sum \bar{Y}_i$$

$$= \frac{41.245}{10} = 4.1245$$

$$V(\bar{Y}) = \left(\frac{1}{n} - \frac{1}{N}\right)S_b^2 + \frac{1}{n}\left(\frac{1}{m} - \frac{1}{M}\right)S_w^2$$

$$S_b^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\bar{Y}_{i.} - \bar{Y}\right)^2$$

$$= \frac{1}{9}(0.7882)$$

$$S_b^2 = 0.0875$$

$$S_b^2 = \frac{1}{n(m-1)}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(Y_{ij} - \bar{Y}_{i.}\right)^2$$

$$= \frac{1}{10(4-1)}(9.5532) = 0.3184$$

$$V(\bar{Y}) = \left(\frac{1}{n} - \frac{1}{N}\right)S_b^2 + \frac{1}{n}\left(\frac{1}{m} - \frac{1}{M}\right)S_w^2$$

$$= \left(\frac{1}{10} - \frac{1}{100}\right)(0.0875) + \frac{1}{10}\left(\frac{3}{16}\right)(0.3184)$$

$$= 0.0078 + 0.0059$$

$$V(\overline{Y}) = 0.0137$$

S.E of $\overline{Y} = \sqrt{V(\overline{Y})} = \sqrt{0.137}$

= 0.1170

ii) In simple random sampling the estimate of variance is given by :

$$V_{ran}(\underline{Y}) = \left(\frac{1}{nm} - \frac{1}{NM}\right)S^2$$

$$S^2 = \frac{1}{NM-1}\left[M(N-1)S^2{}_b + \left\{N(M-1) - (M-m)\frac{N-1}{m}\right\}S^2{}_W\right]$$

$$= \frac{1}{100(16)-1}\left[16(99)(0.0875) + \left\{100(15) - (16-4)\frac{99}{4}\right\}(0.3184)\right]$$

$$= \frac{1}{1600-1}[138.6 + 383.0352]$$

$$= \frac{1}{1599}(521.6352)$$

$$S^2 = 0.3262$$

Then, $V_{ran}(\underline{Y}) = \left[\frac{1}{10(4)} - \frac{1}{100(16)}\right] \times (0.3262)$

= (0.025 - 0.0006) (0.3262)

$$V_{ran}(\overline{Y}) = 0.0079$$

iii) The given cost function is of the form

$C = C_1 n + C_2 mn$ with $C_1 = 4, C_2 = 1, C = 100$

The optimum value of 'm' is given by

$$m_{opt} = \left[\frac{C_1 S_W{}^2}{C_2 S_b{}^2 - \frac{S_W{}^2}{m}}\right]^{1/2}$$

$$= \left[\frac{4(0.3184)}{1(0.0875) - \frac{(0.3184)}{4}}\right]^{1/2}$$

$$= \left[4 \times \frac{0.3184)}{0.0079}\right]^{1/2}$$

$$= \left[4 \times (40.3037)\right]^{1/2}$$

$$m_{opt} = 12.6970$$

Substitute the value of m in given cost function.

The optimum value of n is given by;

$100 = 4n + nm_{opt}$

$= 4n + n\ (12.6970)$

$= 16.6970\ n$

$$n = \frac{100}{16.6970}$$

$= 5.9890$

$$\therefore n_{opt} = 5.9890$$

Conclusion :

i) $\overline{Y} = 4.1245$

$S^2{}_b = 0.0875$

$S^2{}_W = 0.3184$

$V(\overline{Y}) = 0.0137$

ii) $V_{ran}(\overline{Y}) = 0.0079$

iii) Value of $m_{opt} = 12.6970$

$n_{opt} = 5.9890$

PRACTICAL – 10

Estimate the Intra Class Correlation Coefficient

Number of standards of paper in 15 cluster of 4 fields each, selected by simple random sampling without replacement out of 300 fields.

i) Estimate the average number of standard per field along with its standard error.

ii) Estimate the intra class correlation coefficient between, fields belonging to the same clusters with respect to simple random sampling without replacement.

Cluster	1st field	2nd field	3rd field	4th field
1	22	18	27	28
2	53	47	38	29
3	43	29	37	47
4	50	47	41	51
5	73	62	58	47
6	65	71	69	59
7	71	75	31	21
8	24	49	43	75
9	21	72	47	72
10	36	43	51	39
11	72	49	56	69
12	68	64	76	57
13	59	72	67	76
14	43	35	71	40
15	76	58	47	34

Aim :

To estimate the average number of standard per field along with its standard errors and to estimate the intraclass correlation coefficient between fields belonging to the same clusters with respect to simple random sampling without replacement.

Procedure :

$$Y_i = \frac{1}{M}\sum_{j=1}^{M} Y_{ij}$$

$$S^2{}_{wi} = \frac{\sum_{j=1}^{M}\left(Y_{ij} - \overline{Y}_i\right)^2}{M-1}$$

$$S^2{}_b = \frac{\sum_{j=1}^{n}\left(Y_i - \overline{\overline{Y}}\right)^2}{n-1}$$

$$\overline{\overline{Y}} = \frac{\sum_{i=1}^{n}\overline{Y_i}}{n}$$

$$S^2 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\left(Y_{ij} - \overline{\overline{Y}}\right)^2}{(nm-1)}$$

$$V(Y) = \frac{(1-f)S_b^2}{n}$$

standard error = $\sqrt{V(\overline{\overline{y}})}$

$$= \sqrt{\frac{1-f}{n}.S_b^2}$$

$$e = \frac{S^2}{M.S^2{}_b}$$

$$\hat{P} = \frac{1-e}{(m-1)e}$$

Calculation :

S.No.	1st field	2nd field	3rd field	4th field	$\overline{Y_i}$	$\left(\overline{Y}_i - \overline{\overline{Y}}\right)^2$	$\sum_{j=1}^{4}\left(Y_{ij} - \overline{\overline{Y}}\right)^2$
1	22	18	27	28	23.75	751.6699	3071.4296
2	53	47	38	29	41.75	88.6723	685.7092
3	43	29	37	47	39	148.0261	776.1044
4	50	47	41	51	47.25	15.3397	122.1088
5	73	62	58	47	60	78.0289	658.1156
6	65	71	69	59	66	220.0297	964.1188
7	71	75	31	21	49.5	2.7775	2278.11

S.No.	1st field	2nd field	3rd field	4th field	$\bar{Y}_i$	$\left(\bar{Y}_i - \bar{\bar{Y}}\right)^2$	$\sum_{j=1}^{4}\left(Y_{ij} - \bar{\bar{Y}}\right)^2$
8	24	49	43	75	47.75	11.6731	1377.4424
9	21	72	47	72	53	3.3613	1795.4452
10	36	43	51	39	42.25	79.5057	444.7728
11	72	49	56	69	61.5	106.7791	780.1164
12	68	64	76	57	66.25	227.5089	1098.7856
13	59	72	67	76	68.50	300.4467	1362.7866
14	43	35	41	40	47.25	15.3397	846.1088
15	76	58	47	34	53.75	6.6739	975.4456
					$\sum \bar{Y}_i =$ 767.5	$\sum_{i=1}^{15}\left(\bar{Y}_i - \bar{\bar{Y}}\right)^2 =$ 2055.8325	$\sum_{i=1}^{15}\sum_{j=1}^{4}\left(Y_{ij} - \bar{\bar{Y}}\right)^2 =$ 17236.5998

$$\bar{\bar{Y}} = \frac{\sum_{i=1}^{15} \bar{Y}_i}{n}$$

$$= \frac{1}{15}[767.5]$$

$$\bar{\bar{Y}} = 51.1666$$

$$S_b^{\,2} = \frac{\sum_{i=1}^{15}\left(\bar{Y}_i - \bar{\bar{Y}}\right)^2}{n-1}$$

$$= \frac{1}{14}(2055.8325)$$

$$S_b^{\,2} = 146.8451$$

$$S^2 = \frac{\sum_{i=4}^{15}\sum_{j=1}^{4}\left(Y_{ij} - \bar{\bar{Y}}\right)^2}{(nm-1)}$$

$$= \frac{17236.5998}{(60-1)}$$

$$= \frac{17236.5998}{59}$$

$S^2 = 292.1457$

$V(\overline{\overline{Y}}) = \frac{(1-f)}{n} . S_b^{\ 2}$

$= \frac{1}{15}\left(1 - \frac{15}{60}\right)(146.8451)$

$V(\overline{\overline{Y}}) = 7.3422$

Standard Erro = $\sqrt{7.3422} = 2.7096$

Relative Efficiency $\Rightarrow e = \frac{S^2}{MS_b^{\ 2}} = \frac{292.1457}{4(146.8451)}$

e = 0.4973

Intra Class Correlation Coefficient

$\hat{\imath}_e = \frac{1-e}{(M-1)e} = \frac{1-(0.4973)}{(3)(0.4973)}$

$\hat{P}_e = 0.3369$

Conclusion :

The estimated average no. of standard per field :

$\overline{\overline{Y}} = 51.1666$

$V(\overline{\overline{Y}}) = 0.6118$

$S_b^{\ 2} = 146.8451$

$S^2 = 292.1457$

Standard Error = 0.7821

Relative Efficiency = 0.4973

Intraclass Correlation Coefficient is

$\hat{P}_e = 0.3369$