JOURNAL
OF THE
INDIAN SOCIETY
OF
AGRICULTURAL STATISTICS

भारतीय कृषि सांख्यिकी संस्था की पत्रिका

ISAS

Vol. 74	No. 1	APRIL 2020

Indian Society of Agricultural Statistics

Founded: January 03, 1947

The Indian Society of Agricultural Statistics is a scientific society. Its aims are to promote the study of and research in Statistical Theory in the widest sense of the term and its applications to Agriculture, Animal Husbandry, Agricultural Economics and allied fields. The Society has broadened its coverage by including research in Computer Applications also into its gamete. Its membership is open to all persons and institutions interested in the aims of the Society. The membership consists of Honorary Members, Patrons, Life Members, Annual Members and Institutional Members. The membership fee is ₹ 500.00 per annum (foreign, US $ 125) for Annual Members and ₹ 5,000.00 (foreign, US $ 350) for Life Members. Membership fee includes subscription to the Journal. The Permanent Institutional Membership fee is ₹ 25,000.00 (foreign, US $ 1,500). Additional information about the Society and application forms for membership may be obtained from the Secretary, Indian Society of Agricultural Statistics, ICAR-IASRI Campus, Library Avenue, Pusa, New Delhi - 110 012 or may be downloaded from the website of the Society at www.isas.org.in

Journal of Indian Society of Agricultural Statistics

The Society also publishes an International Peer Reviewed Journal called "Journal of the Indian Society of Agricultural Statistics" with ISSN 0019-6363. Three issues of the Journal (April, August and December) are published annually. The first volume of the Journal was released in 1948. The Journal devoted to the publication of original research papers on all aspects of Statistics and Computer Applications preferably with innovative applications in Agricultural Sciences or that have a potential application in Agricultural Sciences. The review articles of the topics of current interest are welcome. The Journal also accepts books, monographs and periodicals for review. Special issues on thematic areas of both national and international importance are also brought out. The Journal has a very strong Editorial Board comprising of Chief Editor, Coordinating Editors and Associate Editors.

Abstracted/Indexed in: *Mathematical Reviews, Zentralblatt MATH, Statistical Theory & Method Abstracts, Current Index to Statistics, CABI Abstracts.*

Manuscript Submission: Manuscript should be prepared as per style of the latest issue of the Journal available at: www.isas.org.in. The manuscript may be submitted for a possible publication through e-mail at: editor.isas@gmail.com

Review Process: Journal of the Indian Society of Agricultural Statistics is peer reviewed journal. The paper is assigned a manuscript number immediately on manuscript submission. An associate editor is then assigned for handling the review process of the paper. The editorial process is expected to complete within 36 weeks of its submission.

Note: The contents of a paper published in this Journal are the sole responsibility of the author or authors, and its publication does not imply the concurrence of the Executive Council of the Indian Society of Agricultural Statistics or its Editors.

Typeset and printed by: Academic and Law Serials Publications, 23/4760-61, Ansari Road, Darya Ganj, New Delhi - 110002 (INDIA)

Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 74(1) 2020

CONTENTS

Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 74(1) 2020 1–10

Stratified Randomized Response Model for Multiple Responses

Raghunath Arnab and D.K. Shangodoyin

University of Botswana, Gaborone, Botswana

Received 10 April 2019; Revised 16 August 2019; Accepted 26 August 2019

SUMMARY

Randomized response (RR) techniques are used to collect data on sensitive characteristics. Abdelfatah and Mazloum (2015) extended Odumade and Singh's (2009) RR techniques based on two decks of cards for stratified sampling and claim, on the basis of empirical studies, that their proposed estimators performs better that the Odumade and Singh (2009) RR technique in most situations. In this paper we have proposed alternative estimators for each of the Odumade and Singh (2009), Abdelfatah *et al.* (2011) and Abdelfatah and Mazloum (2015) RR techniques for stratified sampling. The proposed alternative estimators are found be more efficient than the existing estimators. Apart from increased efficiencies, the proposed estimators possess simpler expressions for the estimators of proportion, variances and unbiased estimators of variances.

Keywords: Optimum allocation, Randomized response; Relative efficiency, Sensitive characteristics.

1. INTRODUCTION

While collecting information, directly from respondents, relating to sensitive issues such as induced abortion, drug addiction, duration of suffering from Aids and so on, the respondents very often report untrue values or even refuse to respond (Arnab and Singh (2013)). Warner (1965) introduced an ingenious technique known as randomized response technique (RR) where a respondent supplies indirect response instead of direct response. Thus the RR technique protects privacy of the respondents while increase quality of data by reducing major sources of bias originated from evasive answers and non-responses. Warner's (1965) technique was modified by several researchers e.g. Horvitz et al. (1967), Greenberg *et al.* (1969), Kim (1978), Franklin (1989), Arcos *et al.* (2015) and Rueda *et al.* (2015) among others which increased cooperation of the respondents and improved efficiencies of the estimators. Applications of the RR techniques in real life surveys were reported by Greenberg *et al.* (1969): Illegitimacy of offspring; Abernathy *et al.* (1970): Incidence of induced abortions; Van der Heijden *et al.* (1998): Social security fraud, and Arnab and Mothupi (2015): Sexual habits of University students. Further details are given by Chaudhuri and Mukherjee (1988), Singh (2003) and Arnab (2017) among others.

Recently Abdelfatah and Mazloum (2015) extended Odumade and Singh (2009) and Abdelfatah *et al.* (2011)'s RR techniques to stratified sampling for estimating the proportion π of a certain sensitive characteristic of a population. On the basis of empirical studies Abdelfatah and Reda Mazloum (2015) showed that the Abdelfatah *et al.*'s (2011) RR technique performed better in about 22% of the cases than the Odumade and Singh (2009)'s RR technique when extended to stratified sampling. We will refer Odumade and Singh (2009), Abdelfatah *et al.* (2011) and Abdelfatah and Mazloum (2015) to this paper as OS, AF and AFM respectively.

In this paper, we have proposed alternative estimators for AFM and OS RR models for stratified sampling. The proposed alternative estimator for AFM model has been proven theoretically superior to the existing AF and AFM estimators. It is shown empirically that the proposed alternative estimator perform always better than the OS estimator. The proposed estimators, their variances and unbiased estimators of the variances

Corresponding author: Raghunath Arnab
E-mail address: arnabr@mopipi.ub.bw

are much simpler than the existing AF, AFM and OS estimators.

2. ALTERNATIVE ESTIMATORS FOR OS AND AF RR MODELS

2.1 OS RR technique

In OS RR technique, a sample of size n units is selected from a population by simple random sampling with replacement (SRSWR) method. Each of the selected respondents in the sample is asked to select one card at random from each of the decks: Deck 1 and Deck 2. Each of the decks consists of two types of cards written "I belong to the sensitive group A" and "I do not belong to the sensitive group A" with proportions P and T respectively. The respondent answers "Yes" or "No" if the statement matches his status with the statement written on the card (Arnab *et al.* (2017)).

Deck 1		Deck 2
I belong to the sensitive group A with proportion P		I belong to the sensitive group A with proportion T
I do not belong to the sensitive group A with proportion $1-P$		I do not belong to the sensitive group A with proportion $1-T$

For example if a respondent selects a card written "I belong to the sensitive group A" from the Deck-1 and selects the other card written "I do not belong to the sensitive group A" from the Deck-2, then he/she will supply with a response "Yes, No" if he/she belong to the sensitive group A. On the other hand if the respondent do not belongs to the group A, he/she will supply "No, Yes" as his/her response (Arnab and Shangodoyin (2015)).

Let $n_{11}(\theta_{11})$, $n_{10}(\theta_{10})$, $n_{01}(\theta_{01})$ and $_{00}(\ _{00})$ denote respectively the frequencies (probabilities) of responses (Yes, Yes), (Yes, No), (No, Yes) and (No, No).

Response from Deck 1	Response from Deck 2		Total
	Yes	No	
Yes	n_{11}	n_{10}	$n_{1\bullet}$
No	n_{01}	n_{00}	$n_{0\bullet}$
Total	$n_{\bullet 1}$	$n_{\bullet 0}$	n

2.1.1 Odumade and Singh's Estimator

Odumade and Singh (2009) proposed an unbiased estimator for the population proportion π by minimizing a distance function

$$D=\frac{1}{2}\sum_{i=0}^{1}\sum_{j=0}^{1}\left(\theta_{ij}-n_{ij}/n\right)^2$$

as

$$\hat{\pi}_{os}=\frac{1}{2}+\frac{(P+T-1)(n_{11}-n_{00})(P-T)(n_{10}-n_{00})}{2n[(P+T-1)^2+(P-T)^2]} \quad (2.1)$$

The variance of $\hat{\pi}_{os}$ and an unbiased estimator of the variance of $\hat{\pi}_{os}$ were given respectively as

$$V\left(\hat{\pi}_{os}\right)=\frac{\begin{array}{c}(P+T-1)^2\{PT+(1-P)(1-T)\}+\\(P-T)^2\{T(1-P)+P(1-T)\}\end{array}}{4n[(P+T-1)^2+(P-T)^2]^2}-\frac{(2\pi-1)^2}{4n}$$

$$=\frac{\pi(1-\pi)}{n}+\frac{1}{4n}\left[\frac{\begin{array}{c}(P+T-1)^2\{PT+(1-P)(1-T)\}+\\(P-T)^2\{T(1-P)+P(1-T)\}\end{array}}{[(P+T-1)^2+(P-T)^2]^2}-1\right] \quad (2.2)$$

and

$$\hat{V}\left(\hat{\pi}_{os}\right)=\frac{1}{4(n-1)}\left[\frac{\begin{array}{c}(P+T-1)^2\{PT+(1-P)(1-T)\}+\\(P-T)^2\{T(1-P)+P(1-T)\}\end{array}}{[(P+T-1)^2+(P-T)^2]^2}-(2\hat{\pi}-1)^2\right] \quad (2.3)$$

2.1.2 Jayraj *et al.* (2016) Estimator

Jayaraj *et al.* (2016) proposed an alternative estimator of π by minimizing a weighted distance function

$$D=\frac{1}{2}\sum_{i=0}^{1}\sum_{j=0}^{1}w_{ij}\left(\theta_{ij}-n_{ij}/n\right)^2$$

with $w_{00}=\frac{(1-P)(1-T)}{(1-P-T)}$, $w_{01}=\frac{(1-P)T}{(T-P)}$, $w_{10}=\frac{P(1-T)}{(P-T)}$ and $w_{11}=\frac{PT}{P+T-1}$

as

$$\hat{\pi}_J=\frac{\begin{array}{c}PTn_{11}+P(1-T)n_{10}+(1-P)Tn_{01}+(1-P)(1-T)n_{00}\\-4PT(1-P)(1-T)\end{array}}{n\{1-2P(1-P)-2T(1-T)\}} \quad (2.4)$$

They obtained the variance of $\hat{\pi}_J$ and an unbiased estimator of the variance of $\hat{\pi}_J$ as

$$V(\hat{\pi}_J)=\frac{\pi(1-\pi)}{n}+\frac{(2P-1)^2(2T-1)^2\{P(1-P)+T(1-T)\}\pi}{n\{1-2P(1-P)-2T(1-T)\}^2}+\frac{PT(1-P)(1-T)\{1-16PT(1-P)(1-T)\}}{n\{1-2P(1-P)-2T(1-T)\}^2} \tag{2.5}$$

and

$$\hat{V}(\hat{\pi}_J)=\frac{\pi_J(1-\pi_J)}{n-1}+\frac{(2P-1)^2(2T-1)^2\{P(1-P)+T(1-T)\}\hat{\pi}_J}{n\{1-2P(1-P)-2T(1-T)\}^2}+\frac{PT(1-P)(1-T)\{1-16PT(1-P)(1-T)\}}{n\{1-2P(1-P)-2T(1-T)\}^2} \tag{2.6}$$

3. ALTERNATIVE ESTIMATOR FOR OS MODEL

Let y_i be the value of the sensitive characteristic (say) of the study variable y for the ith respondent (unit) of the population U of size N. Let $y_i=1$ if the respondent possesses the characteristic A and $y_i=0$ otherwise. Then the proportion of the respondents possess the characteristic A in the population is

$$\pi=\frac{1}{N}\sum_{i\in U}y_i \tag{3.1}$$

Let

$$z_i(j)=\begin{cases}1 \text{ if the ith respondent of the jth deck answers "Yes"}\\ 0 \text{ if the ith respondent of the jth deck answers "No"}\end{cases}$$

for $j=1,2$.

Then,

$$E_R\{z_i(1)\}=y_iP+(1-y_i)(1-P),\ V_R\{z_i(1)\}=P(1-P),$$
$$E_R\{z_i(2)\}=y_iT+(1-y_i)(1-T) \text{ and}$$
$$V_R\{z_i(2)\}=T(1-T) \tag{3.2}$$

where E_R and V_R denote respective the expectation and variance with respect to the RR technique.

From the above Eq. (3.2), we find that

$$r_i(1)=\frac{z_i(1)-P(1-P)}{2P-1} \text{ and } r_i(2)=\frac{z_i(2)-T(1-T)}{2T-1} \tag{3.3}$$

satisfy

$E_R\{r_i(1)\}=E_R\{r_i(2)\}=y_i$, $V_R\{r_i(1)\}=\dfrac{P(1-P)}{(2P-1)^2}=\phi_1$, $V_R\{r_i(2)\}=\dfrac{T(1-T)}{(2T-1)^2}=\phi_2$ and the covariance $C_R\{r_i(1),r_i(2)\}=0$ (as the cards are selected independently) (3.4)

From (3.3) and (3.4), we obtain unbiased estimators of π based on the answers from Deck1 and Deck 2 cards respectively as follows:

$$\hat{\pi}_1=\frac{1}{n}\sum_{i=1}^{n}r_i(1)=\frac{\hat{\lambda}_1-(1-P)}{(2P-1)} \text{ and}$$
$$\hat{\pi}_2=\frac{1}{n}\sum_{i=1}^{n}r_i(1)=\frac{\hat{\lambda}_2-(1-T)}{(2T-1)} \tag{3.5}$$

where $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are the proportion of "Yes" answers from the sampled respondents based on the Deck 1 and Deck 2 cards respectively.

Let E_p, V_p and C_p be operators for expectation, variance and covariance with respect to the sampling design p, we have the following estimators:

$$E(\hat{\pi}_1)=E_p\left[\frac{1}{n}\sum_{i=1}^{n}E_R\{r_i(1)\}\right]=E_p\left(\frac{1}{n}\sum_{i=1}^{n}y_i\right)=\pi, \tag{3.6}$$

$$V(\hat{\pi}_1)=E_p\left[\frac{1}{n^2}\sum_{i=1}^{n}V_R\{r_i(1)\}\right]+V_p\left[\frac{1}{n}\sum_{i=1}^{n}E_R\{r_i(1)\}\right]=E_p[\phi_1/n]+V_p\left(\frac{1}{n}\sum_{i=1}^{n}y_i\right)=\frac{\pi(\quad\pi)}{n}+\frac{\quad}{n} \tag{3.7}$$

Similarly,

$$E(\hat{\pi}_2)=\pi \text{ and } V(\hat{\pi}_2)=\frac{\pi(1-\pi)}{n}+\frac{\phi_2}{n} \tag{3.8}$$

Further,

$$\mathrm{Cov}(\hat{\pi}_1,\hat{\pi}_2)=\mathrm{C}_p\left[E_R(\hat{\pi}_1),E_R(\hat{\pi}_2)\right]+E_p\left[\mathrm{C}_R(\hat{\pi}_1,\hat{\pi}_2)\right]=\mathrm{C}_p\left[\left(\frac{1}{n}\sum_{i=1}^{n}y_i\right),\left(\frac{1}{n}\sum_{i=1}^{n}y_i\right)\right]=\frac{\pi(1-\pi)}{n} \tag{3.9}$$

This study propose the following theorem:

Theorem 3.1.

(i) The optimum estimator π based on the linear combination of $\hat{\pi}_1$ and $\hat{\pi}_2$ is

$$\hat{\pi}_{w0}=w_0\hat{\pi}_1+(1-w_0)\hat{\pi}_2$$

where $w_0 = \dfrac{\phi_2}{\phi_1 + \phi_2}$

(ii) The variance of $\hat{\pi}_{w0}$ is

$$V\left(\hat{\pi}_{w0}\right) = \frac{\pi(1-\pi)}{n} + \frac{\bar{\phi}}{n}$$

where

$$\bar{\phi} = \left(\frac{1}{\phi_1} + \frac{1}{\phi_2}\right)^{-1} = \frac{P(1-P)T(1-T)}{P(1-P)\left(2T-1\right)^2 + T(1-T)\left(2P-1\right)^2}$$

(iii) An unbiased estimator of $V\left(\hat{\pi}_{w0}\right)$ is

$$\hat{V}\left(\hat{\pi}_{w0}\right) = \frac{\hat{\pi}_{w0}\left(1-\hat{\pi}_{w0}\right) + \bar{\phi}}{n-1}$$

Proof:

Consider an unbiased estimator of π based on the linear combination of $\hat{\pi}_1$ and $\hat{\pi}_2$ as

$$\hat{\pi}_w = w\hat{\pi}_1 + \left(1-w\right)\hat{\pi}_2 \tag{3.10}$$

The variance of $\hat{\pi}_w$ is

$$V\left(\hat{\pi}_w\right) = V_p\left[E_R\left(\hat{\pi}_w\right)\right] + E_p\left[V_R\left(\hat{\pi}_w\right)\right]$$

$$= V_p\left[\frac{1}{n}\sum_{i=1}^{n} y_i\right] + E_p\left[w^2\phi_1 + \left(1-w\right)^2\phi_2\right]/n$$

$$= \frac{\pi\left(1-\pi\right)}{n} + \frac{w^2\phi_1 + \left(1-w\right)^2\phi_2}{n} \tag{3.11}$$

Differentiating $V\left(\hat{\pi}_w\right)$ with respect to w and then equating it to zero, the optimum value of w comes out as $w_0 = \dfrac{\phi_2}{\phi_1 + \phi_2}$.

(ii) Substituting $w = w_0 = \dfrac{\phi_2}{\phi_1 + \phi_2}$ in (3.11), we find

$$V\left(\hat{\pi}_{w0}\right) = \frac{\pi(1-\pi)}{n} + \frac{\bar{\phi}}{n}$$

(iii) $$E\left[\hat{V}\left(\hat{\pi}_{w0}\right)\right] = \frac{E\left[\hat{\pi}_{w0}\left(1-\hat{\pi}_{w0}\right)\right] + \bar{\phi}}{n-1}$$

$$= \frac{E\left(\hat{\pi}_{w0}\right) - E\left(\hat{\pi}_{w0}\right)^2 + \bar{\phi}}{n-1}$$

$$= \frac{\pi\left(1-\pi\right) + \bar{\phi} - V\left(\hat{\pi}_{w0}\right)}{n-1}$$

$$= V\left(\hat{\pi}_{w0}\right)$$

4. EFFICIENCY OF THE PROPOSED ESTIMATOR

The percentage relative efficiency of Jayraj *et al.* (2016) estimator $\hat{\pi}_J$ and the proposed estimator $\hat{\pi}_{w0}$ compared with Odumade and Singh (2009) estimator $\hat{\pi}_{os}$ are given by

$$E\left(1\right) = \frac{V\left(\hat{\pi}_{os}\right)}{V\left(\hat{\pi}_J\right)} \times 100 \tag{4.1}$$

and

$$E\left(2\right) = \frac{V\left(\hat{\pi}_{os}\right)}{V\left(\hat{\pi}_{w0}\right)} \times 100 \tag{4.2}$$

It is important to note that the differences $V\left(\hat{\pi}_J\right) - V\left(\hat{\pi}_{os}\right)$ and $V\left(\hat{\pi}_J\right) - V\left(\hat{\pi}_{wo}\right)$ increase with π. Hence Jayraj *et al.* (2016) estimator performs worse than $V\left(\hat{\pi}_{os}\right)$ and $V\left(\hat{\pi}_{wo}\right)$ for higher values of π. The relative percentage efficiencies $E\left(1\right)$ and $E\left(2\right)$ are given in the following Table 4.1 for different values of P, T and π. The Table 4.1 shows that the proposed estimator $\hat{\pi}_{wo}$ perform better than $\hat{\pi}_{os}$ in all situations while $\hat{\pi}_{wo}$ performs better than $\hat{\pi}_J$ in most of situations. Both the estimators $\hat{\pi}_{wo}$ and $\hat{\pi}_{os}$ perform better than $\hat{\pi}_J$ for higher values of π in general. However for (P, T)=(0.1, 0.4), (0.2, 0.4) and (0.3, 0.4), $\hat{\pi}_J$ perform better than the other two while for (P, T)=(0.2, 0.1), (0.3, 0.1), (0.3, 0.2), (0.4, 01) and (0.4, 0.2), $\hat{\pi}_J$ performs very poor.

Remark 4.1

Jayaraj *et al.* (2017) proposed an alternative estimator for the proportion π and found empirically that their proposed estimator is superior to the Odumade and Singh (2009) estimator. The proposed Jayaraj *et al.*'s (2017) will be subject of our future investigation.

5. AF RR MODELS

Under AF model, each respondent is asked to draw two cards; one from the "Deck 1" and another from "Deck 2". Deck 1 comprises two types of cards as in Warner (1965) model viz. "I belong to the sensitive group A" with proportion P and "I do not belong to the sensitive group A" with proportion $1-P$. The respondent should answer truthfully "Yes" if the statement matches his/her status otherwise, answers "No". The Deck 2 comprises also two types of cards written "Yes" with proportion Q and "No" with proportion 1-Q. Regardless of his/her actual status the

Table 4.1: Relative efficiencies E(1) and (E2)

π	P = 0.1							
	T= 0.1		T= 0.2		T= 0.3		T= 0.4	
	E(1)	E(2)	E(1)	E(2)	E(1)	E(2)	E(1)	E(2)
0.1	128.0	100	156.5	104.3	179.3	107.1	185.9	103.5
0.2	108.0	100	127.9	103.1	146.0	105.4	153.7	102.7
0.3	99.7	100	116.5	102.6	133.2	104.6	141.7	102.3
0.4	94.3	100	109.9	102.4	126.6	104.2	136.3	102.1
0.5	89.7	100	105.1	102.3	122.8	104.1	134.1	102.0
0.6	85.0	100	101.1	102.4	120.6	104.2	134.3	102.1
0.7	79.4	100	96.9	102.6	119.6	104.6	137.2	102.3
0.8	71.7	100	91.9	103.1	119.9	105.4	144.7	102.7
0.9	59.5	100	84.2	104.3	122.6	107.1	164.5	103.5
π	P = 0.2							
	T= 0.1		T= 0.2		T= 0.3		T= 0.4	
	E(1)	E(2)	E(1)	E(2)	E(1)	E(2)	E(1)	E(2)
0.1	93.3	104	130.8	100	188.8	101.2	244.4	101.2
0.2	82.3	103	112.2	100	159.0	101.1	204.8	101.1
0.3	75.8	103	102.3	100	144.3	101.0	186.6	101.0
0.4	70.6	102	95.4	100	135.6	100.9	177.4	100.9
0.5	65.5	102	89.7	100	129.9	100.9	173.4	100.9
0.6	60.0	102	84.3	100	125.9	100.9	173.3	100.9
0.7	53.6	103	78.5	100	123.0	101.0	177.3	101.0
0.8	45.8	103	71.8	100	120.9	101.1	187.2	101.1
0.9	35.7	104	63.1	100	119.4	101.2	208.8	101.2
π	P = 0 .3							
	T= 0.1		T= 0.2		T= 0.3		T= 0.4	
	E(1)	E(2)	E(1)	E(2)	E(1)	E(2)	E(1)	E(2)
0.1	32.1	107	54.8	101.2	116.8	100	242.5	100.3
0.2	33.1	105	53.3	101.1	108.7	100	220.0	100.2
0.3	32.5	105	51.3	101.0	102.7	100	206.4	100.2
0.4	31.0	104	48.9	100.9	97.9	100	198.2	100.2
0.5	28.8	104	46.1	100.9	93.7	100	193.5	100.2
0.6	25.9	104	42.9	100.9	89.7	100	191.7	100.2
0.7	22.5	105	39.1	101.0	85.7	100	192.6	100.2
0.8	18.5	105	34.7	101.1	81.4	100	196.5	100.2
0.9	13.8	107	29.5	101.2	76.5	100	204.4	100.3
π	P=0.4							
	T= 0.1		T= 0.2		T= 0.3		T= 0.4	
	E(2)	E(1)	E(2)	E(1)	E(2)	E(1)	E(2)	E(1)
0.1	2.5	104	5.3	101.2	18.4	100.3	104.4	100
0.2	3.0	103	5.7	101.1	18.7	100.2	102.7	100
0.3	3.3	102	6.0	101.0	18.8	100.2	101.1	100
0.4	3.4	102	6.0	100.9	18.7	100.2	99.6	100
0.5	3.4	102	5.9	100.9	18.3	100.2	98.1	100
0.6	3.1	102	5.7	100.9	17.8	100.2	96.7	100
0.7	2.8	102	5.3	101.0	17.0	100.2	95.3	100
0.8	2.3	103	4.7	101.1	16.1	100.2	93.9	100
0.9	1.7	104	4.1	101.2	15.0	100.3	92.4	100

respondent has to answer "Yes" if he /she receives card written "Yes". Alternatively, if the respondent receives the card written "No" the respondent should answer "No" as his or her response.

Deck 1		Deck 2
I $\in A$ with proportion P		"Yes" with proportion Q
I $\in A^c$ with proportion $1-P$		"No" with proportion $1-Q$

Let the responses of the selected sample of n units by SRSWR method be classified as follows:

Response from Deck 1	Response from Deck 2		Total
	Yes	No	
Yes	n_{11}	n_{10}	$n_{1\cdot}$
No	n_{01}	n_{00}	$n_{0\cdot}$
Total	$n_{\cdot 1}$	$n_{\cdot 0}$	n

By using the above scenario, AF (2011) derived the following results:

(i) An unbiased estimator of the population π is

$$\hat{\pi}_f = \frac{1}{2} + \frac{Q(n_{11}/n - n_{01}/n) + (1-Q)(n_{10}/n - n_{00}/n)}{2(2P-1)[Q^2 + (1-Q)^2]},$$

$$P \neq 0.5 \tag{5.1}$$

(ii) The variance of $\hat{\pi}_f$ is

$$V(\hat{\pi}_f) = \frac{Q^3 + (1-Q)^3}{4n(2P-1)^2[Q^2 + (1-Q)^2]^2} - \frac{(2\pi - 1)^2}{4n},$$

$P \neq 0.5$

$$= \frac{\pi(1-\pi)}{n} + \frac{1}{4n}\left[\frac{Q^3 + (1-Q)^3}{(2P-1)^2[Q^2 + (1-Q)^2]^2} - 1\right] \tag{5.2}$$

(iii) An unbiased estimator of the variance of $\hat{\pi}_f$ is

$$\hat{V}(\hat{\pi}_f) = \frac{1}{4(n-1)}\left[\frac{Q^3 + (1-Q)^3}{(2P-1)^2[Q^2 + (1-Q)^2]^2} - (2\hat{\pi}_f - 1)^2\right] \tag{5.3}$$

5.1 Improved estimator for AF model

AF argued that the force responses "Yes" or "No" will increase respondents' confidentiality and cooperation as the respondents need not answer sensitive question twice as in OS model. Since, the response of the Deck 2 has no relation with the sensitive characteristic of the respondents (variable under study), Arnab and Singh (2013) recommended that one should ignore response of the Deck 2 for the analysis of the data. So, they proposed a modified estimator based on the responses from the Deck 1 only. The proposed alternative estimator is

$$\hat{\pi}_1 = \frac{\hat{\lambda}_1 - (1-P)}{2P-1} \tag{5.4}$$

where $\hat{\lambda}_1$ is the proportion of "Yes" answers from the Deck 1.

The properties of the estimator $\hat{\pi}_1$ are obtained from Chaudhuri and Mukherjee (1988) as follows:

(i) $(\hat{\pi}\ )\quad \pi$

(ii) $V(\hat{\pi}_1) = \frac{\pi(1-\pi)}{n} + \frac{P(1-P)}{n(2P-1)^2}$

(iii) An unbiased estimator of $V(\hat{\pi}_1)$ is

$$\hat{V}(\hat{\pi}_1) = \frac{\hat{\lambda}_1(1-\hat{\lambda}_1)}{n(2P-1)^2}$$

6. STRATIFIED SAMPLING

Consider a finite population stratified into H strata. Let N_h be the total number of units and π_h be the proportion of individuals possess the sensitive characteristic A in the stratum h. Then, $\pi = \sum_{h=1}^{h} N_h \pi_h / N$ be the proportion of individuals possessing the characteristic A in the entire population of size $N = \sum_{h=1}^{H} N_h$. From each of the stratum samples are selected by SRSWR method independently. Let n_h be the number of respondents selected from the hth stratum.

In this section we will compare performances of the alternative estimators for OS and AF methods of RR techniques when extended to the stratified sampling. The RR techniques for the stratified samplings is described as follows:

6.1 OS model

Each of the respondents of the selected sample of the hth stratum is asked to draw one card from each of the two decks independently with proportions $P_h(\neq 1/2)$ and $T_h(\neq 1/2)$. The details of the cards and data obtained from the stratum h are as follows:

Stratum h

Deck 1		Deck 2
I $\in A$ with proportion P_h		I $\in A$ with proportion T_h
I $\in A^c$ with proportion $1-P_h$		I $\in A^c$ with proportion $1-T_h$

Responses obtained from stratum h

Response from Deck 1	Response from Deck 2		Total
	Yes	No	
Yes	$n_{11}(h)$	$n_{10}(h)$	$n_{1\cdot}(h)$
No	$n_{01}(h)$	$n_{00}(h)$	$n_{0\cdot}(h)$
Total	$n_{\cdot 1}(h)$	$n_{\cdot 0}(h)$	n_h

OS estimator of π for the stratified sampling was proposed by AFM as follows:

$$T_{os} = \frac{1}{N}\sum_{h=1}^{H} N_h \hat{\pi}_{os}^h \tag{6.1}$$

where

$$\hat{\pi}_{os}^h = \frac{1}{2} + \frac{(P_h + T_h - 1)(n_{11}(h) - n_{00}(h))(P_h - T_h)(n_{10}(h) - n_{00}(h))}{2n_h[(P_h + T_h - 1)^2 + (P_h - T_h)^2]}.$$

Proposed alternative estimator for stratified sampling based on OS RR technique is

$$T_{ws} = \frac{1}{N}\sum_{h=1}^{H} N_h \hat{\pi}_{wo}^h \tag{6.2}$$

where

$\hat{\pi}_{wo}^h = w_o^h \hat{\pi}_1^h + \left(1 - w_o^h\right)\hat{\pi}_2^h$, $\hat{\pi}_1^h = \frac{\hat{\lambda}_1^h - (1-P_h)}{(2P_h - 1)}$, $\hat{\pi}_2^h = \frac{\hat{\lambda}_2^h - (1-T_h)}{(2T_h - 1)}$, $\hat{\lambda}_1^h\left(\hat{\lambda}_2^h\right)$ = proportion of "Yes" answers from the Deck 1 (Deck 2), $w_0^h = \frac{\phi_2^h}{\phi_1^h + \phi_2^h}$, $\phi_1^h = \frac{P_h(1-P_h)}{(2P_h-1)^2}$, $\phi_2^h = \frac{T_h(1-T_h)}{(2T_h-1)^2}$.

The variances of T_{os} and T_{ws} are given by

$$V(T_{os}) = \frac{1}{N^2}\sum_{h=1}^{H} N_h^2 \frac{\sigma_{hos}^2}{n_h} \text{ and}$$

$$V\left(T_{ws}\right)=\frac{1}{N^2}\sum_{h=1}^{H}N_h^2\frac{\sigma_{hws}^2}{n_h} \tag{6.3}$$

where

$$\sigma_{hos}^2=\pi_h\left(1-\pi_h\right)+\frac{1}{4}\left[\frac{\begin{array}{l}(P_h+T_h-1)^2\{P_hT_h+(1-P_h)(1-T_h)\}+\\(P_h-T_h)^2\{T_h(1-P_h)+P_h(1-T_h)\}\end{array}}{[(P_h+T_h-1)^2+(P_h-T_h)^2]^2}-1\right],$$

$$\sigma_{hws}^2=\pi_h(1-\pi_h)+\bar{\phi}_h \text{ and } \bar{\phi}_h=\left(\frac{1}{\phi_1^h}+\frac{1}{\phi_2^h}\right)^{-1}.$$

6.2 AF model

In this model also each respondents of the stratum h is asked to draw one card at random from each of two decks independently with proportions $W_h\left(\neq 1/2\right)$ and $Q_h\left(\neq 1/2\right)$ respectively. Here the respondent matches his/her status with the statement written on the card drawn from the Deck-1 and answers "Yes" or "No". For the card drawn from the Deck-2, respondents answer "Yes" or "No" on the basis of "Yes" or "No" written in the card.

Stratum h

Deck 1		Deck 2
I $\in A$ with proportion W_h		"Yes" with proportion Q_h
I $\in A^c$ with proportion $1-W_h$		"No" with proportion $1-Q_h$

Responses obtained from stratum h

Response from Deck 1	Response from Deck 2		Total
	Yes	No	
Yes	$n_{11}(h)$	$n_{10}(h)$	$n_{1\cdot}(h)$
No	$n_{01}(h)$	$n_{00}(h)$	$n_{0\cdot}(h)$
Total	$n_{\cdot 1}(h)$	$n_{\cdot 0}(h)$	n_h

Using this scenario, Abdelfatah and Mazloum (2015) proposed the following estimator for the population proportion π .

$$T_f=\frac{1}{N}\sum_{h=1}^{H}N_h\hat{\pi}_f^h \tag{6.4}$$

where

$$\hat{\pi}_f^h=\frac{1}{2}+\frac{Q_h\left(n_{11}(h)-n_{01}(h)\right)+(1-Q_h)\left(n_{10}(h)-n_{00}(h)\right)}{2(2W_h-1)[Q_h^{\;2}+(1-Q_h)^2]n_h}$$

The proposed alternative estimator for AFM is

$$T_{1f}=\frac{1}{N}\sum_{h=1}^{H}N_h\hat{\pi}_{1f}^h \tag{6.5}$$

where $\hat{\pi}_{1f}^h=\dfrac{\hat{\lambda}_{1h}-W_h(1-W_h)}{2W_h-1}$ and $\hat{\lambda}_{h1}$ = Proportion of "Yes" answers from Deck 1 of hth stratum for AFM RR technique.

The variances of T_f and T_{1f} are as follows:

$$V\left(T_f\right)=\frac{1}{N^2}\sum_{h=1}^{H}N_h^2\frac{\sigma_{hf}^2}{n_h} \tag{6.6}$$

$$V\left(T_{1f}\right)=\frac{1}{N^2}\sum_{h=1}^{H}N_h^2\frac{\sigma_{1hf}^2}{n_h} \tag{6.7}$$

where

$$\sigma_{hf}^2=\pi_h(1-\pi_h)+\frac{1}{4}\left[\frac{Q_h^3+(1-Q_h)^3}{(2W_h-1)^2[Q_h^2+(1-Q_h)^2]^2}-1\right]$$

and $\sigma_{1hf}^2=\pi_h\left(1-\pi_h\right)+\dfrac{W_h(1-W_h)}{\left(2W_h-1\right)^2}$.

6.3 Optimum allocation

Consider the simple cost function for stratified sampling suggest by Cochran (1977) as

$$C=c_o+\sum_{h=1}^{H}c_hn_h \tag{6.8}$$

where c_0 is the overhead fixed cost and c_h is the cost per unit for the hth stratum.

The optimum sample sizes n_h that minimizes the variance of the form

$$\Psi=\frac{1}{N^2}\sum_{h=1}^{H}N_h^2\frac{\sigma_h^2}{n_h} \tag{6.9}$$

keeping the total cost of the survey fixed as C^* is given by

$$n_{h0}=\frac{C^*-c_0}{\sum_{h=1}^{H}N_h\sigma_h\sqrt{c_h}}\frac{N_h\sigma_h}{\sqrt{c_h}} \tag{6.10}$$

The optimum value of Ψ with $n_h=n_{ho}$ is

$$\Psi_0=\frac{1}{N^2\left(C^*-c_0\right)}\left(\sum_{h=1}^{H}N_h\sigma_h\sqrt{c_h}\right)^2 \tag{6.11}$$

For the Neyman allocation $c_h=c$ and the total sample size $n=\sum_h n_h=\left(C^*-c_0\right)/c$ is fixed. In this case Ψ_0 in (6.11) reduces to

$$\Psi_{ney}=\frac{1}{N^2}\left(\sum_{h=1}^{H}N_h\sigma_h\right)^2/n \tag{6.12}$$

The expressions of the variances under Neyman allocation for the estimators T_{0s}, T_{ws}, T_f and T_{1f} are respectively given by

$$\mathrm{V}_{os} = \frac{1}{n}\left(\sum_{h=1}^{H} Z_h \sigma_{hos}\right)^2, \quad \mathrm{V}_{ws} = \frac{1}{n}\left(\sum_{h=1}^{H} Z_h \sigma_{hws}\right)^2,$$

$$\mathrm{V}_f = \frac{1}{n}\left(\sum_{h=1}^{H} Z_h \sigma_{hf}\right)^2 \text{ and } \mathrm{V}_{1f} = \frac{1}{n}\left(\sum_{h=1}^{H} Z_h \sigma_{1hf}\right)^2 \quad (6.13)$$

where $Z_h = N_h / N$.

6.4 Efficiency Comparison

For the AFM RR model, the proposed alternative estimator T_{1f} is more efficient than T_f as $\sigma_{1hf} \le \sigma_{hf}$. The modified estimator T_{ws} for OS strategy with $P_h = W_h$ is more efficient than T_{1f} as $\sigma_{1hf} \ge \sigma_{hws}$. Following Abdelfatah and Mazloum (2015), we compare relative percentage efficiencies of the estimators T_{0s}, T_{ws}, T_{1f} with respect to T_f numerically and these are given in Table 6.1 for $h = 2$, $P_1 = P_2 = P$; $T_1 = T_2 = T$; $W_1 = W_2 = W$; $Q_1 = Q_2 = Q$ and different combination of Z_h, π_{1h} and π_{2h} as follows: $P(=W) = 0.1, 0.2, 0.3, 0.4$, $T(=Q) = 0.1, 0.2, 0.3, 0.4$, $Z_1(=1-Z_2) = 0.1, 0.3, 0.5, \; 0.7, 0.9$ and $(\pi_1, \pi_2) =$ (0.08, 0.13), (0.38, 0.53), (0.78, 0.83), (0.85, 0.95). The relative percentage efficiencies of T_{0s}, T_{ws}, T_{1f} with respect to T_f are given by

$$EOS = \frac{V_f}{V_{os}} \times 100, \; E1 = \frac{V_f}{V_{1f}} \times 100 \text{ and } EW = \frac{V_f}{V_{ws}} \times 100$$

The empirical studies reveal that the estimator T_{ws} performs the best in all the situations. The next place is occupied by T_{0s}. The improved estimator T_{1f} is more efficient than T_f but less efficient than T_{0s}. However, the comparison between T_{1f} and T_{0s} is not fair as the estimator T_{1f} is based on the responses of Deck 1 cards only while T_{0s} is based on the responses of both Deck 1 and Deck 2 cards.

Table 6.1. Relative efficiencies of the estimators T_{0s}, T_{1f} and T_{ws} with respect to T_f

P(=W)	T(=Q)	$Z_1(=1-Z_2) = 0.1$											
		$\pi_1 = .08,\ \pi_2 = 0.13$			$\pi_1 = 0.38,\ \pi_2 = 0.53$			$\pi_1 = 0.78,\ \pi_2 = 0.83$			$\pi_1 = 0.85,\ \pi_2 = 0.95$		
		EOS	E1	EW	EOS	E1	EW	EOS	E1	EW	EOS	E1	EW
0.1	0.1	133	119	138	133	109	133	148	112	148	182	117	182
	0.2	115	116	122	120	113	123	129	117	133	143	125	151
	0.3	104	106	107	110	110	114	113	114	119	119	120	129
	0.4	273	111	284	102	104	104	103	105	106	105	107	109
0.2	0.1	193	116	193	207	109	212	250	110	258	328	112	345
	0.2	140	113	141	166	112	166	184	115	184	211	117	211
	0.3	109	104	111	130	110	131	137	112	138	145	114	147
	0.4	618	109	659	107	104	108	109	104	110	110	105	112
0.3	0.1	362	114	366	434	109	452	555	109	587	759	110	823
	0.2	206	111	206	300	112	303	343	113	347	398	114	403
	0.3	124	104	125	190	110	190	201	111	201	214	111	214
	0.4	2611	109	2695	122	104	122	124	104	124	125	104	126
0.4	0.1	1302	113	1317	1726	109	1761	2307	109	2372	3282	109	3417
	0.2	566	110	568	1052	112	1062	1227	113	1241	1437	113	1456
	0.3	204	104	204	518	110	519	553	110	554	588	110	590
	0.4	133	119	138	199	104	199	202	104	202	205	104	205

P(=W)	T(=Q)	$Z_1(=1-Z_2) = 0.3$											
		$\pi_1 = .08,\ \pi_2 = 0.13$			$\pi_1 = 0.38,\ \pi_2 = 0.53$			$\pi_1 = 0.78,\ \pi_2 = 0.83$			$\pi_1 = 0.85,\ \pi_2 = 0.95$		
		EOS	E1	EW	EOS	E1	EW	EOS	E1	EW	EOS	E1	EW
0.1	0.1	160	114	160	133	109	133	147	112	147	173	116	173
	0.2	134	120	140	121	113	123	128	117	132	140	123	146
	0.3	115	116	123	110	110	114	113	113	119	117	118	127
	0.4	104	106	107	102	104	104	103	105	106	104	107	108
0.2	0.1	280	111	291	208	109	213	246	110	254	308	112	323
	0.2	195	116	195	166	113	166	183	115	183	205	117	205
	0.3	140	113	142	130	110	132	136	112	138	143	113	145
	0.4	109	105	111	107	104	108	109	104	110	110	105	111

0.3	0.1	635	109	678	436	109	455	545	109	576	709	110	765
	0.2	367	114	371	301	112	304	340	113	344	386	114	391
	0.3	207	111	207	190	110	190	201	111	201	212	111	212
	0.4	124	104	125	122	104	122	123	104	124	125	104	125
0.4	0.1	2691	109	2781	1737	109	1773	2262	109	2324	3047	109	3163
	0.2	1320	113	1336	1056	112	1066	1215	113	1228	1394	113	1412
	0.3	569	110	571	518	110	520	551	110	552	581	110	583
	0.4	204	104	204	199	104	199	202	104	202	205	104	205

P(=W)	T(=Q)	$Z_1(=1-Z_2)=0.5$											
		$\pi_1=.08,\ \pi_2=0.13$			$\pi_1=0.38,\ \pi_2=0.53$			$\pi_1=0.78,\ \pi_2=0.83$			$\pi_1=0.85,\ \pi_2=0.95$		
		EOS	E1	EW	EOS	E1	EW	EOS	E1	EW	EOS	E1	EW
0.1	0.1	163	114	163	133	109	133	146	111	146	166	115	166
	0.2	136	121	141	121	113	124	127	116	131	137	121	142
	0.3	116	117	124	110	110	114	112	113	119	116	117	125
	0.4	104	106	107	102	104	104	103	105	106	104	106	108
0.2	0.1	286	111	298	209	109	214	243	110	251	292	111	304
	0.2	198	116	198	166	113	166	182	114	182	200	116	200
	0.3	141	113	143	130	110	132	136	112	137	142	113	143
	0.4	110	105	111	107	104	108	109	104	110	110	105	111
0.3	0.1	653	110	699	439	109	457	536	109	566	667	110	715
	0.2	372	114	376	302	113	305	337	113	341	375	114	380
	0.3	208	111	208	190	110	190	200	111	200	209	111	209
	.4	125	104	125	122	104	122	123	104	124	125	104	125
0.4	0.1	2777	109	2873	1748	109	1785	2218	109	2278	2845	109	2945
	0.2	1339	113	1355	1060	112	1070	1203	113	1216	1353	113	1370
	0.3	573	110	574	519	110	520	548	110	550	575	110	576
	0.4	204	104	204	199	104	199	202	104	202	204	104	204

P(=W)	T(=Q)	$Z_1(=1-Z_2)=0.7$											
		$\pi_1=.08,\ \pi_2=0.13$			$\pi_1=0.38,\ \pi_2=0.53$			$\pi_1=0.78,\ \pi_2=0.83$			$\pi_1=0.85,\ \pi_2=0.95$		
		EOS	E1	EW	EOS	E1	EW	EOS	E1	EW	EOS	E1	EW
0.1	0.1	167	115	167	133	109	133	145	111	145	160	114	160
	0.2	137	122	143	121	113	124	127	116	131	134	120	139
	0.3	116	117	125	110	110	114	112	113	118	115	116	123
	0.4	104	106	108	102	104	105	103	105	106	104	106	107
0.2	0.1	294	111	307	210	109	215	240	110	247	278	111	289
	0.2	200	116	200	167	113	167	180	114	180	195	116	195
	0.3	142	113	144	131	110	132	135	111	137	140	113	142
	0.4	110	105	111	107	104	108	108	104	110	109	104	111
0.3	0.1	672	110	721	441	109	460	528	109	556	630	109	672
	0.2	377	114	381	303	113	306	334	113	338	365	114	370
	0.3	210	111	210	191	110	191	199	111	199	207	111	207
	0.4	125	104	125	122	104	122	123	104	123	124	104	125
0.4	0.1	2869	109	2971	1760	109	1797	2176	109	2234	2668	109	2756
	0.2	1358	113	1375	1064	112	1074	1192	113	1205	1315	113	1330
	0.3	576	110	577	520	110	521	546	110	548	568	110	570
	0.4	204	104	204	199	104	199	202	104	202	204	104	204

P(=W)	T(=Q)	$Z_1(=1-Z_2)=0.9$											
		$\pi_1=.08,\ \pi_2=0.13$			$\pi_1=0.38,\ \pi_2=0.53$			$\pi_1=0.78,\ \pi_2=0.83$			$\pi_1=0.85,\ \pi_2=0.95$		
		EOS	E1	EW	EOS	E1	EW	EOS	E1	EW	EOS	E1	EW
0.1	0.1	170	115	170	134	109	134	143	111	143	155	113	155
	0.2	138	122	145	121	113	124	126	116	130	132	119	136
	0.3	117	118	126	110	110	114	112	113	118	114	115	121
	0.4	104	106	108	102	104	105	103	104	106	103	105	107
0.2	0.1	302	111	315	210	109	216	237	110	244	266	111	275
	0.2	203	117	203	167	113	167	179	114	179	190	115	190
	0.3	143	113	144	131	110	132	135	111	136	139	112	140
	0.4	110	105	111	107	104	108	108	104	109	109	104	110

0.3	0.1	693	110	745	444	109	462	519	109	546	598	109	635
	0.2	382	114	387	304	113	307	331	113	335	356	114	360
	0.3	211	111	211	191	110	191	198	111	198	205	111	205
	0.4	125	104	125	122	104	122	123	104	123	124	104	124
0.4	0.1	2967	109	3076	1772	109	1809	2136	109	2191	2513	109	2590
	0.2	1378	113	1395	1068	112	1078	1180	113	1193	1279	113	1293
	0.3	579	110	581	521	110	522	544	110	545	562	110	564
	0.4	205	104	205	200	104	200	202	104	202	203	104	203

7. CONCLUSION

An alternative estimator $\hat{\pi}_{wo}$ for OS RR model has been proposed. The proposed estimator perform better than OS estimator $\hat{\pi}_{os}$ always while it performs better than the estimator $\hat{\pi}_J$ most of the situations. Abdelfatah *et al.* (2011) and Abdelfatah and Mazloum (2015) used RR techniques based two decks of cards where the Deck 1 relates to sensitive questions and Deck 2 relates to non-sensitive questions. They anticipated that their proposed RR techniques increase level of confidentiality and hence co-operation from the respondents. They showed empirically that their proposed RR strategy for stratified sampling performs better than Odumade and Singh (2009) RR model. In this paper, we have shown that Abdelfatah *et al.* (2011) and Abdelfatah and Mazloum (2015) estimators can always be improved by using information of sensitive question on Deck 1 card and ignoring responses for the unrelated questions based on the card 2. Table 6.1 shows that the alternative estimator T_{ws} always perform better than Odumade and Singh (2009) estimator. It is also worth noting that the both the proposed estimator possess very simple expression for the estimator of the proportions, variances and unbiased estimator of variances.

ACKNOWLEDGEMENTS

The authors are thankful to the anonymous referee for his valuable suggestions regarding improvement of the paper.

REFERENCES

Abdelfatah, S. and Mazloum, R. (2015). An improved stratified randomized response model using two decks of cards, *Model Assisted Statistics and Applications*, **10**, 309-320.

Abdelfatah, S., Mazloum, R. and Singh, S. (2011). An alternative randomized response model using two decks of cards, *Statistica*, **71**, 381-390

Abernathy, J.R., Greenberg, B.G. and Horvitz D.G. (1970). Estimates of induced abortion in urban North Carolina, *Demography*, **7**, 19-29.

Arcos, A., Rueda, M., and Singh, S.A. (2015). Generalized approach to randomized response for quantitative variables, *Quality & Quantity*, **49**, 1239-1256.

Arnab, R. (2017). *Survey Sampling Theory and Applications*, Academic Press, U.K.

Arnab, R. and Mothupi, T. (2015). Randomized response techniques: A case study of the risky behaviors' of students of a certain University, *Model Assisted Statistics and Applications,* **10**, 421-430

Arnab, R. and Shangodoyin, D.K. (2015). Randomized response techniques using maximum likelihood method, *Communications in Statistics-Theory and Methods*, **44**, 3340-3349.

Arnab, R., Shangodoyin, D.K.. and Kgosi, P.M. (2017). Randomized response techniques with multiple responses. *J. Ind. Soc. Agril. Statist*., **71**, 201-205.

Arnab, R. and Singh, S. (2013). An alternative randomized response model using two deck of cards: A rejoinder. *Statistica,* **LXXIII**, 493-503.

Chaudhuri, A. and Mukherjee, R. (1988). *Randomized response: Theory and Techniques*, New York, Marcel Dekker.

Cochran, W.G. (1977). *Sampling Techniques*, 3rd Ed. New York : Wiley.

Franklin, L.A., (1989). A comparison of estimators for randomized response sampling with continuous distribution from dichotomous populations. *Communication in Statistics- Theory and methods,* **18**, 489-505.

Greenberg, B.D., Abul-Ela, A.L.A., Simmons, W.R., and Horvitz, D.G., (1969). The unrelated question randomized response model: Theoretical framework. *J. Amer. Statist. Assoc.,* **64**, 520-539.

Horvitz, D.G., Shah, B.V., and Simmons, W.R., (1967). The unrelated question randomized response model. Proceedings of Social Statistics Section, *American Statistical Association*., **65**-72.

Jayaraj, A., Odumade, O., Sedory, S. and Singh, S. (2016). Weighted squared distance for two- decks randomized response model, *Joint Statistical meeting, Survey research section, American Statistical Association,* 1014-1028.

Jayaraj, A., Sedory, S., Odumade, O., and Singh, S. (2017). A new estimatior of proportion with linear function using data from two decks randomized response model. *Communication in Statistics-Theory and methods,DOI:10.1080/03610926.2017.1321764*

Kim, J.I. (1978). Randomized response technique for surveying human populations. Ph.D. Dissertation, Temple University, Philadelphia, USA.

Odumade, O. and Singh, S. (2009). Efficient use of two deck of cards in randomized response sampling. *Communication in Statistics-Theory and methods,* **38**, 439-446.

Rueda, M., Cobo, B., Arcos, A., (2015). Package 'RRTCS': Randomized Response Techniques for Complex Surveys. http://cran.r-project.org/web/packages/RRTCS/.

Singh, S. (2003). *Advanced Sampling Theory with Applications*, Kluwer Academic Publishers, Netherlands.

Van der Heijden, P.G.M., Van Gils, G., Bouts, J., and Hox, J. J. (1998). A comparison of randomized response. CASAQ, and direct questioning; eliciting sensitive information in the context of social security fraud. *Kwantitatieve Methoden,* **59**, 15-34.

Warner, S. L., (1965). Randomize response: a survey technique for eliminating evasive answer bias. *J. Amer. Statist. Assoc.,* **60**, 63-69.

Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 74(1) 2020 11–16

Climate Resilient, High Yielding and Stable Sugarcane Genotypes in India

Rajesh Kumar[1], A.D. Pathak[1] and Bakshi Ram[2]

[1]*ICAR-Indian Institute of Sugarcane Research, Lucknow*
[2]*ICAR-Sugarcane Breeding Institute, Caimbatore*

Received 28 June 2019; Revised 09 October 2019; Accepted 17 October 2019

SUMMARY

Based on long term data analysis of advanced varietal trials of All India Coordinated Research Project (AICRP) on Sugarcane, identified nine genotypes which possess the qualities of high yield sustainability and low sensitivity towards adverse changes in environmental conditions during 2012 to 2018. Out of nine, two Co 10024 and Co 11001 from early group and two CoM 11086 and Co 08009 from mid late group of Peninsular Zone. In East Coast Zone, only one early maturing genotype, CoA 13322, was identified. Two mid-late genotypes, CoH 08262 and CoH 09264, were identified from North West Zone. Similarly two mid-late genotypes, CoSe 11454, CoP 12438, were identified from North Central and North Eastern Zone. Out of 163 genotypes, only CoSe 11454 was highly stable for all the three character, CCS (t/ha), cane yield (t/ha) and sucrose (%). Other eight were highly stable for CCS (t/ha) and cane yield (t/ha) only. These may be used as parents in crossing programme as possess the qualities of high yield sustainability and low sensitivity towards adverse changes in environmental conditions or having high stability and high yield criteria.

Keywords: Climate resilient, Sugarcane, Genotypes, Sustainability, Stability.

1. INTRODUCTION

About 38% districts (241 of 634) in India were found to be resilient to drought to dry condition (Annon., 2018). In fact, deficit monsoon has become chronic with 13 of the last 18 years witnessing below-normal rains. In recent years, country had continuously five deficit monsoon since 2014. Sugarcane is a high biomass crop which requires large amounts of water for good yields. Irrigation is necessary in order to produce sugarcane in almost all parts of the country, but water supplies are becoming increasingly limited (Gupta and Kumar, 2018). Most of the sugarcane area is resilient and slightly non-resilient to drought to dry condition and this crop shows better tolerance to water extreme than other crops. That is why the sugarcane yield in the country was not effected due to drought to dry condition and resulted sugarcane productivity in between 65 to 70 t/ha in last twenty years except 2002-03, 2015-16 and 2016-17. These three years were had either normal or little deficit monsoon. As sugarcane is highly sensitive to climatic and edaphic factors, location specific selection of varieties is important, as varietal requirement differs for every zone (Anon, 2014). This study was under taken to identify sugarcane genotypes which possess the qualities of high yield sustainability and low sensitivity towards adverse changes in environmental conditions.

2. MATERIAL AND METHODS

Genotype x Environment (GE) interaction continues to be a challenging issue among plant breeders, geneticists and agronomists in conducting varietal trials across diverse environments. Methods of partitioning GE interaction into components measure the contribution of each genotype in GE interaction. Whenever an interaction is significant, use of main effects e.g., overall genotype means across environments is often questionable. Stability performance of genotype is considered as an important aspect in varietal trials. Researchers need a statistics that provides a reliable measure of stability or consistency of performance of a

Corresponding author: Rajesh Kumar
E-mail address: rajesh_iisr@hotmail.com

genotype across a range of environments, particularly one that reflects the contribution of each genotype to the total GE interaction and helps in identifying the best genotype. For a successful breeding or genotype testing programme, both stability and yield (or any other trait) must be simultaneously considered. Also integration of stability of performance with yield through suitable measures will help in selecting genotypes in a more precise manner. In this study, it is proposed to use simultaneous selection indices using Additive Main Effects and Multiplicative Interaction (AMMI) model. This model is appropriate when main effects (genotypic, environmental) and genotype x environment interaction (GE) effects are both important in yield trials.

AMMI model offers a more appropriate statistical analysis to deal with such situations, compared to traditional methods like ANOVA, Principal Component Analysis (PCA) and linear regression. Currently, selection of sugarcane genotypes is based on the performance of cane yield across the location in a zone and ranking of genotypes is done on the basis of mean data. Ranking of genotypes based on simultaneous selection of high yielding and stable genotypes gives better and reliable picture in identifying a variety for release.

Simultaneous selection approach proposed by Rao & Prabhakaran (2005) and Kumar & Sinha (2012 to & 2015) was used in this study which selects genotypes for both high yield and stability in multi-environmental trials using AMMI model by assigning 80% weight to yield and 20% to stability values of the genotype.

2.1 AMMI and simultaneous selection procedure

The AMMI method combines the traditional ANOVA and PCA into a single analysis with both additive and multiplicative parameters (Gauch, 1992). The first part of AMMI uses the normal ANOVA procedures to estimate the genotype and environment main effects. The second part involves the PCA of the interaction residuals (residuals after main effects are removed). The model formulation for AMMI shows its interaction part consists of summed orthogonal products. Because of this form the interaction lends itself to graphical display in the form of so-called bi-plots (Gabriel (1971)). Here, it is assumed that the first two PCA axes suffice for an adequate description of the GxE interaction. It is evident from earlier sections that the scope of bi-plots is very much limited. The inferences drawn from bi-plots will be valid only when the first two PCAs explain a large portion of interaction variation. In situations, where more than two PCA axes are needed to accumulate considerable portion of GEl variation, what should be the approach for identifying varieties which are high yielding as well as stable. Keeping this in mind, a new family of simultaneous selection indices was proposed by Rao and Prabhakaran (2005) which can select varieties for both yield and stability was applied in this study. The proposed selection indices (I_i) consists of (i) a yield component, measured as the ratio of the average performance ($\bar{Y}_{i.}$) of the i-th genotype to the overall mean performance of the genotypes under test, and (ii) a stability component, measured as the ratio of stability information (l/ ASTAB$_i$) of the i-th genotype to the mean stability information of all the genotypes under test. The simultaneous selection index is given as

$$I_i = \frac{\bar{Y}_{i.}}{\bar{Y}} + a\frac{\frac{1}{ASTAB_i}}{\frac{1}{T}\sum_{i=1}^{T}\frac{1}{ASTAB_i}}$$

Where ASTAB$_i$ is as stability measure of the i-th genotype under AMMI procedure and $\bar{Y}_{i.}$ is mean performance of i-th genotype. α is the ratio of the weights given to the stability components (w_2) and yield (w_1) with a restriction that $w_1 + w_2 = 1$.

Simultaneous selection criterion proposed by Rao and Prabhakaran (2005) was used in this study which selects genotypes for both high yield and stability in multi-environmental trials using AMMI model by assigning 80 % weight to yield and 20 % to stability value of the genotypes. Such weights were assigned because Hogart (1976) inferred that 75 % of the gains in cane yield in Australia were attributed to the varietal improvement and Edme *et al.* (2005) estimated that genetic improvement along contributed 69 % of the sugarcane yield.

This method was used for selection of high yielding and stable genotypes under Advance Varietal Trial of early and midlate maturity group in Plant I & II and ratoon crops conducted during (2012 to 2018) in Peninsular Zone, East Coast Zone, North West Zone and North Central Zone (Map 1) of All India Coordinated Project on Sugarcane. Advance Varietal Trials (Plant I) were conducted during first year and same crop was ratooned during second year of the crop. Advance Varietal Trials (Plant II) were conducted during second year of the trial. Combination of two years of plant crops and one ratoon crop data were analyzed for stability analysis. AMMI analyses and simultaneous selection indices analyses were performed with the help of SAS 9.3 (SAS Institute, 2002-2010). Other statistical analysis was done using Ms-Excel (2014). In each zone, ranking of varieties was based on the above mentioned criterion for commercial cane sugar (CCS t/ha), cane yield (t/ha) and sucrose (%). Similar analysis was done for each identified genotype (Table 1) with other genotype of the trial for Simultaneous selection criterion proposed by Rao and Prabhakaran (2005).

2.2 New initiative of genotype ranking against high yield sustainability and low sensitivity towards adverse changes in environmental conditions under crop improvement programme of AICRP(S)

A successful evaluation of genotypes for stable performance under varying environmental conditions based on information on genotype × environment interaction for yield is an essential part of any sugarcane varietal development programme. The selection of sugarcane genotypes is based on the performance of cane yield across the location in a zone and ranking of genotypes was done on the basis of mean data. The same criterion was used in All India Coordinated Research Project (AICRP) on Sugarcane since 1971 and till 2011-12. A new approach involving simultaneous selection indices using Additive Main Effects and Multiplicative Interaction (AMMI) model for Advanced Varietal Trial of All India Coordinated Research Project on Sugarcane has been applied for simultaneous selection of high yielding and stable sugarcane genotypes. The approach involves three steps for selection of high yielding and stable genotype, in Advanced Varietal Trial of AICRP on sugarcane. In the first step, genotypes performing better than the best standards in the trial based on only yield performance are selected. In second step, the selected genotypes are ranked / judged on index values obtained on basis of both yield and stability. The third step involves the ranking of selected genotypes of step one on basis of their stability. Genotypes are considered best, high yielding and stable, if their respective ranks were found better than the ranks of best standard or at least one of the standards. If their ranks are inferior to the best standard, then we judged the top ranks among the tested genotypes based on index value.

3. RESULTS AND DISCUSSION

Based on the large data analysis of proposed above procedure, out of 163 genotypes tested in five zones at Advanced Varietal Trials under All India Coordinated Research Project (AICRP) on Sugarcane, identified nine genotypes which possess the qualities of high yield sustainability and low sensitivity towards adverse changes in environmental conditions during 2012 to 2018 (Table 1). Because these entries were found high yielding for cane yield (t/ha), CCS (t/ha) and Sucrose (%) and stable as per the procedure suggested by Rao & Prabhakaran (2005) except entry Co 13322 of East Coast Zone (Table 3, 4 and 5). These entries were also highly stable in respective zones if we consider only AMMI stability procedure proposed by Gauch (1992). Co 13322, an early maturing genotype, ranked second in the zone for cane yield (t/ha), CCS (t/ha) and Sucrose (%). But this entry recorded the highest cane yield of 117.91 t/ha among the nine entries. This entry had 13.57 t/ha CCS and 16.55 % sucrose.

Out of nine, two Co 10024 and Co 11001 from early group and two CoM 11086 and Co 08009 from midlate group were highly stable and high cane yielding of Peninsular Zone (Table 3, 4 and 5). These four entries recorded cane yield in between 94 to 106 t/ha. Similarly, these entries were highly stable and high yielding for CCS (t/ha). Among the nine entries, Co 08009 recorded the highest CCS (t/ha) of 14.07 t/ha and 19.35 % sucrose (Table 3, 4 and 5).

In North West Zone, out of nine, two CoH 08262, CoH 09264 from midlate group were highly stable and high cane yielding for cane yield (t/ha) and CCS (t/ha). Both these were developed from CCS HAU research centre, Uchani. These entries had low values of stability for CCS (t/ha) and sucrose (%) and had high rank for cane yield and CCS (t/ha).

In North Central and North Eastern Zone, mid-late entry CoSe 11454 was found to be the only entry during this period which showed the high stability for sucrose (%) along with cane yield (t/ha) and CCS (t/ha) among all the nine entries. Other eight entries had high stability for cane yield and CCS (t/ha) only. For Sucrose (%), these eight entries had very inferior rank and high value of stability among the nine entries. CoSe 11454 had lowest value of stability (2.48), which indicate that it is very stable genotype. This entry recorded cane yield (74.5 t/ha), CCS (9.14 t/ha) and Sucrose (17.6 %). This entry may be considered for release for commercial cultivation in North Central Zone of All

India Coordinated Research Project on Sugarcane. Similar performance were also observed for the mid-late entry, CoP 12438, in this zone.

For sugarcane, if the simultaneous selection index value is around 1.45 then genotype is high yielding and highly stable across the zone for cane yield (t/ha) and CCS (t/ha). Similarly for sucrose (%) if it is around 1.20 then the genotype is high yielding and highly stable across the zone. As far as pest and diseases reaction of the genotypes is concern, the information of these nine genotypes is presented in Table 2, at an advance level of screening. Normally, genotypes are promoted to an advance level of testing which are resistant or moderately resistant to diseases at initial level of screening. Similar situation is for two important pest of sugarcane (Table 2). All are LS - Least susceptible to Early Shoot Borer and Top Borer.

CONCLUSION

Climate change induced changes in growth and development and adverse effects on sugarcane and sugar productivity invoke in urgency for climate – resilient varieties of sugarcane to mitigate such effects. Nine identified genotypes can also be considered as climate resilient genotypes of different zones (Table 1). These genotypes will be least effected by drought and water logging in different part of the country because yield fluctuations were minimum in the trials due to high stability in cane yield. Use of these nine genotypes, as parents, in sugarcane breeding programme may also be helpful in imparting multiple – stress tolerance and sustaining sugarcane production and productivity in different zones of AICRP(S) in the country under such conditions. These genotypes may be considered for release for commercial cultivation in different zone of All India Coordinated Research Project on Sugarcane.

REFERENCES

Anonymous (2014): Principal Investigator’s Report (Varietal Improvement Programme), All India Coordinated Research Project on Sugarcane, Sugarcane Breeding Institute, Coimbatore.

Anonymous (2018): Only 32% of India is resilient to drought. The Hindu October 23.

Gabriel, K.R. (1971): The biplot-graphical display ofmatrices with applications to principal component analysis. *Biometrika* 58, 453-467.

Gauch, H.G. (1992): *Statistical analysis of regional yield trials: AMMI analysis of factorial designs.* Elsevier, Amsterdam.

Gupta, R. and Kumar, R. (2018): Micro Irrigation Systems in Sugarcane in India. Paper presented during 72nd Annual Conference of Indian Society of Agricultural Statistics at ICAR - Central Institute of Agricultural Engineering, Bhopal Madhya Pradesh on December 13-15, 2018

Kumar, Rajesh and Sinha, O.K. (2015): Simultaneous selection of high yielding and stable mid-late maturing sugarcane genotypes of East Coast Zone in India using AMMI Model : A new approach. *Indian Journal of Sugarcane Technology,* 30(01): 19-27.

Kumar, Rajesh and Sinha, O.K. (2016): A new approach of simultaneous selection of high sugar yielding and stable genotypes of East Coast Zone in India using AMMI model. Paper presented and published in the Proceedings of 74th Annual Convention of The Sugar Technologists’ Association of India held from 28-30 July 2016 at Delhi.

Rao, A.R., Prabhakaran, V.T., (2005): Use of AMMI in simultaneous selection of genotypes for yield and stability. J. Ind. Soc. Ag. Stat. 59(1), 76-82.

Acknowledgement

Authors are thankful to Director, ICAR-Sugarcane Breeding Institute, Coimbatore and Principal Investigator (Varietal Improvement Programme), All India Coordinated Research Project on Sugarcane for their support and guidance. Authors are also thankful to Heads/In-charges of centres of All India Coordinated Research Project on Sugarcane for conducting the trials and providing the data for publication in Principal Investigator (Varietal Improvement Programme) report, AICRP(S), ICAR-Sugarcane Breeding Institute, Coimbatore, Tamil Nadu used in the paper.

Table 1. Climate resilient high yielding and stable sugarcane genotypes in India

Zone	Early	Mid-late
Peninsular Zone	Co 10024, Co 11001,	CoM 11086, Co 08009
East Coast Zone	CoA 13322	
North West Zone		CoH 08262, CoH 09264,
North Central and North Eastern Zone		CoSe 11454, CoP 12438

Table 2. Details and performance and of climate resilient high yielding and stable sugarcane genotypes

Genotype	Zone name of the tested genotype	No. of Entries tested in trial	Maturity	Year	Trial Conducted	Disease reaction			Pest reaction	
						*Red rot	YLD	Smut	ESB	TB
Co 10024	Peninsular Zone	11	Early	2017	Plant I, Plant II and Ratoon	R or MR	R	S	LS	-
Co 11001	Peninsular Zone	8	Early	2018	Plant I, Plant II and Ratoon	R or MR	R or MR	R or MR	LS	-
CoM 11086	Peninsular Zone	8	Midlate	2018	Plant I, Plant II and Ratoon	R or MR	R or MR	R or MR	LS	-
Co 08009	Peninsular Zone	7	Midlate	2014	Plant I, Plant II and Ratoon	R or MR	-	R or MR	-	-
CoA 13322	East Coast Zone	7	Early	2018	Plant I, Plant II and Ratoon	R or MR	-	-	LS	-
CoH 08262	North West Zone	9	Midlate	2014	Plant I, Plant II and Ratoon	R or MR	-	S to MR	-	LS
CoH 09264	North West Zone	8	Midlate	2015	Plant I, Plant II and Ratoon	R or MR	R	MS to R	-	LS
CoSe 11454	North Central and North Eastern Zone	7	Midlate	2017	Plant I, Plant II and Ratoon	R or MR	-	R	MS	LS
CoP 12438	North Central and North Eastern Zone	6	Midlate	2018	Plant I, Plant II and Ratoon	R or MR	-	-	LS	LS

*Red rot reaction by plug and cotton swab method at least one centre of the zone
R – Resistant, MR - Moderately Resistant, LS - Least susceptible MS - Moderately susceptible
ESB – Early Shoot Borer and TB – Top Borer

Table 3. Ranking of genotypes of according to their (i) mean performance, (ii) stability and (iii) simultaneous index value in respect of cane yield (t/ha)

Genotype	Maturity	Estimated value			Rank based on estimated value		
		Index Value	Cane Yield (t/ha) value	Stability value	Index value based rank	Cane Yield (t/ha) based rank	Stability based rank
Peninsular Zone							
Co 10024	Early	1.56	98.07	1366.91	1	4	1
Co 11001	Early	1.45	94.51	1275.3	1	2	1
CoM 11086	Midlate	1.52	101.26	681.61	1	2	1
Co 08009	Midlate	1.5	102.06	993.65	1	3	1
East Coast Zone							
CoA 13322	Early	1.37	117.91	238.15	2	1	3
North West Zone							
CoH 08262	Midlate	1.55	81.85	171.71	1	2	1
CoH 09264	Midlate	1.56	85.28	275.66	1	1	1
North Central and North Eastern Zone							
CoSe 11454	Midlate	1.71	74.50	95.88	1	4	1
CoP 12438	Midlate	1.52	76.59	95.96	1	1	1

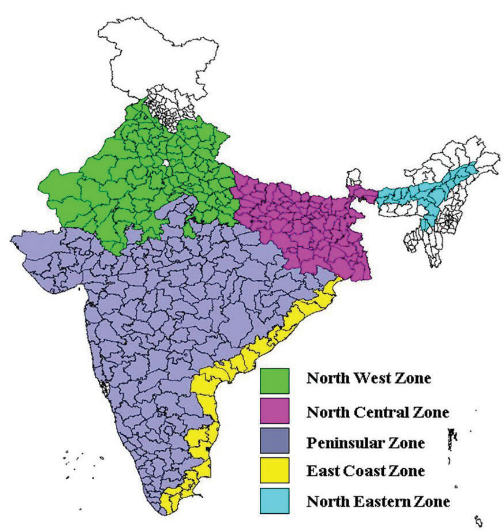
Table 4. Ranking of genotypes of according to their (i) mean performance, (ii) stability and (iii) simultaneous index value in respect of CCS (t/ha)

Genotype	Maturity	Estimated value			Rank based on estimated value		
		Index Value	CCS (t/ha) value	Stability value	Index value based rank	CCS (t/ha) based rank	Stability based rank
Peninsular Zone							
Co 10024	Early	1.51	12.18	25.86	1	5	1
Co 11001	Early	1.37	11.29	22.42	1	3	1

Genotype	Maturity	Estimated value			Rank based on estimated value		
		Index Value	CCS (t/ha) value	Stability value	Index value based rank	CCS (t/ha) based rank	Stability based rank
CoM 11086	Midlate	1.4	13.19	19.22	1	4	1
Co 08009	Midlate	1.43	14.07	20.28	1	3	1
East Coast Zone							
CoA 13322	Early	1.35	13.57	6.41	2	2	3
North West Zone							
CoH 08262	Midlate	1.44	9.65	3.15	1	5	1
CoH 09264	Midlate	1.46	9.62	5.16	1	2	1
North Central and North Eastern Zone							
CoSe 11454	Midlate	1.51	9.14	3.37	1	2	1
CoP 12438	Midlate	1.45	9.09	2.00	1	3	1

Table 5. Ranking of genotypes of according to their (i) mean performance, (ii) stability and (iii) simultaneous index value in respect of Sucrose (%)

Genotype	Maturity	Estimated value			Rank based on estimated value		
		Index Value	Sucrose(%) value	Stability value	Index value based rank	Sucrose (%) based rank	Stability based rank
Peninsular Zone							
Co 10024	Early	1.23	17.56	7.35	7	8	8
Co 11001	Early	1.11	17.02	7.03	8	8	8
CoM 11086	Midlate	1.26	18.53	7.33	3	6	3
Co 08009	Midlate	1.23	19.35	3.98	4	3	4
East Coast Zone							
CoA 13322	Early	1.21	16.55	3.73	7	8	5
North West Zone							
CoH 08262	Midlate	1.26	16.90	2.81	5	7	3
CoH 09264	Midlate	1.11	16.57	5.82	8	8	8
North Central and North Eastern Zone							
CoSe 11454	Midlate	1.33	17.6	2.48	1	2	1
CoP 12438	Midlate	1.32	17.06	2.2	2	4	2



Map 1: Zones of All India Coordinated Research Project on Sugarcane

Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 74(1) 2020 17–22

Some Transformed and Composite Chain Ratio-type Estimators using Two Auxiliary Variables

Sarvesh Kumar Dubey[1], B.V.S. Sisodia[1] and Manoj Kumar Sharma[2]

[1]*Narendra Deva University of Agriculture and Technology, Faizabad*
[2]*Giri Institute of Development Studies, Lucknow*

Received 24 January 2019; Revised 12 September 2019; Accepted 17 October 2019

SUMMARY

The present paper has dealt some transformed and composite chain ratio type estimators of population mean in finite population survey sampling using information on two auxiliary variables related to the variable under study. Their bias and MSE are derived. The relative efficiency of the proposed estimators have been examined in comparison of several existing estimators in the literature using real data. It has been found that the proposed estimators have outperformed the existing chain ratio type estimators and composite chain ratio type estimators.

Keywords: Two phase sampling, Auxiliary variables, Chain ratio type estimators, Composite chain estimators.

1. INTRODUCTION

Information on auxiliary variables is generally used in sample surveys to improve the efficiency of the estimators of population parameter of interest. Theoretically, it has been established that, in general, the regression estimator is more efficient than the ratio and product estimators. However when the regression line of the character under study on the auxiliary character passes through the origin, these are equally efficient. Nevertheless, due to the stronger intuitive appeal, statisticians are more inclined towards the use of ratio and product estimators. Perhaps that is why an extensive work has been done in the direction of improving the performance of these estimators.

Consider that finite population $U = (U_1, U_2, \ldots\ldots U_N)$ consists of N identifiable sampling units. Associated with the unit U_i is a pair of numbers (y_i, x_i), where y_i is the value of the study variate y and x_i is the value of an auxiliary variable x related to y. The objective is to estimate $\bar{Y} = \sum_{i=1}^{N} y_i \big/ N$

For estimating of $\bar{Y}$, let a sample of size n is drawn from U by simple random sampling without replacement (SRSWOR). Let $\bar{y}$ and $\bar{x}$ be simple sample mean of y and x, respectively. When $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} x_i$ is known, the usual ratio estimator of $\bar{Y}$ is given by

$$\bar{y}_r = \frac{\bar{y}}{\bar{x}}.\bar{X} \tag{1.1}$$

with bias and mean square error (MSE) as follows.

$$\mathrm{B}(\bar{y}_r) = .\bar{Y}\left(\frac{1}{n} - \frac{1}{N}\right)\left(C_x^2 - \rho_{yx} C_x C_y\right) \tag{1.2}$$

$$\mathrm{MSE}(\bar{y}_r) = \bar{Y}^2\left(\frac{1}{n} - \frac{1}{N}\right)\left[C_y^2 + C_x^2 - 2\rho_{yx} C_x C_y\right] \tag{1.3}$$

where $C_y^2 = S_y^2 \big/ \bar{Y}^2$, $C_x^2 = S_x^2 \big/ \bar{X}^2$, ρ_{yx} is correlation between x and y, and $S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2$, $S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2$.

The usual regression estimator of $\bar{\mathrm{Y}}$ is given by

Corresponding author: B.V.S. Sisodia
E-mail address: bvssisodia@gmail.com

$$\bar{y}_{lr} = \bar{y} + b_{yx}\left(\bar{X} - \bar{x}\right) \tag{1.4}$$

with bias and MSE as follows.

$$B\left(\bar{y}_r\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\beta\left[\frac{\mu_{30}}{S_x^2} - \frac{\mu_{21}}{S_{xy}}\right] \tag{1.5}$$

$$MSE\left(\bar{y}_r\right) = \left(\frac{1}{n} - \frac{1}{N}\right)S_y^2\left(\; - \; {}_{yx}^{2}\right) \tag{1.6}$$

where $\underset{r}{S_{xy}} = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \bar{X}\right)\left(y_i - \bar{Y}\right)$ and $\mu_{rs} = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \bar{Y}\right)^r\left(x_i - \bar{X}\right)^s$.

When $\bar{X}$ is not known the two phase sampling (double sampling) techinque is resorted to first find out the estimate of $\bar{X}$. The technique involves two steps.

(i) at first phase, draw a preliminary large sample of size n' from U by SRSWOR. Enumerate the sampled units for x. Let $\bar{x}' = \sum_{i=1}^{n'} x_i \Big/ n$ be sample mean based on n' units, and (ii)at second phase, draw a sub sample of size *n* from n' by SRSWOR. Enumerate the sampled unites for *y*. Let $\bar{y}$ and $\bar{x}$ be sample mean based on n units drawn from $'$ units. A double sampling ratio estimator of $\bar{Y}$ is given by

$$\bar{y}_{rd} = \frac{\bar{y}}{\bar{x}}\bar{x}' \tag{1.7}$$

with bias and MSE of $\bar{y}_{rd}$ as given below

$$B\left(\bar{y}_{rd}\right) = \bar{Y}\left(\frac{1}{n} - \frac{1}{n'}\right)\left(C_x^2 - \rho_{yx}C_xC_y\right) \tag{1.8}$$

$$MSE\left(\bar{y}_{rd}\right) = V(\bar{y}) + \bar{Y}^2\left(\frac{1}{n} - \frac{1}{n'}\right)\left[C_x^2 - 2\rho_{yx}C_yC_x\right],$$

$$\text{where } V(\bar{y}) = \left(\frac{1}{n} - \frac{1}{N}\right)S_y^2 \tag{1.9}$$

The usual double sampling regression estimator is given by

$$\bar{y}_1 = \bar{y} + b_{yx}\left(\bar{x}' - \bar{x}\right) \tag{1.10}$$

with bias and MSE as follows

$$B\left(\bar{y}_1\right) = \beta\left(\frac{1}{n} - \frac{1}{n'}\right)\left[\frac{\mu_{30}}{S_{xy}} - \frac{\mu_{21}}{S_{xy}}\right] \tag{1.11}$$

$$MSE\left(\bar{y}_1\right) = V(\bar{y}) - \bar{Y}^2\left(\frac{1}{n} - \frac{1}{n'}\right)\rho_{yx}^2C_y^2 \tag{1.12}$$

Chand (1975) proposed chain ratio-type estimator of $\bar{Y}$ when the population mean of an auxiliary variable *x* highly related with the study variable y is not known but the population mean of another auxiliary variable *z* which is less correlated with *y* but is known. The estimator he proposed is highly correlated with *x* as

$$T_1 = \bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right)\left(\frac{\bar{Z}}{\bar{z}'}\right)$$

$$\text{where } \bar{Z} = \sum_{i=1}^{N} Z_i \Big/ N \text{ and } \bar{z}' = \sum_{i=1}^{n'} z_i \Big/ n' \tag{1.13}$$

and MSE of T_1 as follows

$$MSE\left(T_1\right) = V\left(\bar{y}\right) + \bar{Y}^2\left[\left(\frac{1}{n} - \frac{1}{n'}\right)C_x^2 + \left(\frac{1}{n'} - \frac{1}{N}\right)C_z^2 - 2\left(\frac{1}{n} - \frac{1}{n'}\right)\rho_{yx}C_yC_x - 2\left(\frac{1}{n'} - \frac{1}{N}\right)\rho_{yz}C_yC_z\right] \tag{1.14}$$

Kiregyera (1980) proposed chain ratio-cum regression estimator of $\bar{Y}$ as

$$T_{l2} = \bar{y} + b_{yx}\left(\bar{x}'\frac{\bar{Z}}{\bar{z}'} - \bar{x}\right) \tag{1.15}$$

where b_{yx} is estimated regression coefficient of *y* on *x* from *n* sampled units.

with MSE of T_{l2} of as follows

$$MSE\left(T_{l2}\right) = MSE\left(\bar{y}_1\right) + \bar{Y}^2\frac{f'}{n'}\left[\frac{\rho_{yx}^2C_y^2C_z^2}{C_x^2} - 2\frac{\rho_{yx}\rho_{yz}C_y^2C_z}{C_x}\right],$$

$$\text{where } f' = \left(1 - \frac{n'}{N}\right) \tag{1.16}$$

Kiregyera (1984) developed another estimator of $\bar{Y}$ along with its MSE as

$$T_{l3} = \bar{y} + b_{yx}\left[\bar{x}' + b_{xz}\left(\bar{z} - \bar{z}'\right) - \bar{x}\right] \tag{1.17}$$

$$MSE(T_{l3}) = \bar{Y}^2\left[\left(\frac{1}{n} - \frac{1}{N}\right)\left(C_y^2 - K_{yx}^2C_x^2\right) + \left(\frac{1}{n'} - \frac{1}{N}\right)\left(K_{yx}^2C_x^2 + K_{yx}K_{xz}C_z^2\left(K_{yx}K_{xz} - 2K_{yz}\right)\right)\right] \tag{1.18}$$

where $K_{yx} = \frac{\rho_{yx}C_y}{C_x}$, $K_{xz} = \frac{\rho_{xz}C_x}{C_z}$ and $K_{yz} = \frac{\rho_{yz}C_y}{C_z}$ and b_{xz} is estimated regression coefficient of *x* on *z* from n' sampled units.

Since then many research workers have developed exponential chain ratio type and regression type estimator of $\bar{Y}$ in the past. Some relevant works are briefly described below.

Singh and Choudhury (2012) proposed a exponential chain ratio estimator under double sampling as

$$\bar{Y}_{\mathrm{Re}}^{dc} = \bar{y}\ \exp\left(\frac{\frac{\bar{x}'\bar{z}}{\bar{z}'} - \bar{x}}{\frac{\bar{x}'\bar{z}}{\bar{z}'} + \bar{x}}\right) \tag{1.19}$$

Khare *et al.* (2013) developed a generalized chain ratio cum –regression type estimator of $\bar{Y}$ as

$$T_{l4} = \bar{y} + b_{yx}\left[\bar{x}'\left(\frac{\bar{Z}}{\bar{z}'}\right)^{\alpha} - \bar{x}\right] \tag{1.20}$$

where α is some scalar quantity.

MSE of T_{l4} for optimum value of as $\alpha$is given by

$$MSE\left(T_{l4}\right) = MSE\left(\bar{y}_{lrd}\right) - \bar{Y}^2 \frac{f'}{n'}\rho_{yz}^2 C_y^2 \tag{1.21}$$

Singh *et al.* (2015) developed a composite type chain ratio estimator of $\bar{Y}$ as follows

$$\bar{Y}_{EC}^{RdR} = \left(\alpha I_1 + \left(1-\alpha\right) I_2\right) \tag{1.22}$$

where α is some constant and $I_1 = \bar{Y}_{\mathrm{Re}}^{dc}$ as given in (1.19), and $I_2 = \bar{Y}_{EdR}^{dc} = \bar{y}\ \exp\left(\frac{\frac{N\frac{\bar{x}'\bar{z}}{\bar{z}'} - n\bar{x}}{N-n} - \frac{\bar{x}'\bar{z}}{\bar{z}'}}{\frac{N\frac{\bar{x}'\bar{z}}{\bar{z}'} - n\bar{x}}{N-n} + \frac{\bar{x}'\bar{z}}{\bar{z}'}}\right)$

where I_2 is dual chain ratio estimator of $\bar{Y}$ following the dual to ratio estimator for Srivenkataramana (1977).

MSE of $\bar{Y}_{EC}^{RdR}$ for optimum value of α is given by

$$MSE(\bar{Y}_{EC}^{RdR})_{opt} = \bar{Y}^2\left[\frac{1-f}{n}C_y^2 + N_1 + N_2 - \frac{\left(N_3+N_4\right)^2}{N_5}\right] \tag{1.23}$$

where $N_1 = \frac{\left(1-f^*\right)}{n}\frac{g^2}{4}C_x^2\left(1-\frac{g}{4}K_{yx}\right)$,

$N_2 = \frac{\left(1-f_1\right)}{n_1}\frac{g^2}{4}C_z^2\left(1-\frac{g}{4}K_{yz}\right)$,

$N_3 = \frac{\left(1-f^*\right)}{n}\frac{g}{2}C_x^2\left(1-\frac{2}{g}K_{yx}\right)$,

$N_4 = \frac{\left(1-f_1\right)}{n_1}\frac{g}{2}C_z^2\left(1-\frac{2}{g}K_{yz}\right)$,

$N_5 = \frac{\left(1-f^*\right)}{n}C_x^2 + \frac{\left(1-f_1\right)}{n_1}C_z^2$

and $g = \frac{n}{N-n}$, $f = \frac{n}{N}$, $f_1 = \frac{n'}{N}$, $f^* = \frac{n}{n'}$

MSE of $\bar{Y}_{EdR}^{dc}$ is given by

$$MSE\left(\bar{Y}_{EdR}^{dc}\right) = \bar{Y}^2\left[\left(\frac{1-f}{n}\right)C_y^2 + N_1 + N_2\right] \tag{1.24}$$

In view of the above facts, some transformed and composite chain ratio type estimators of population mean $\bar{Y}$ are proposed using two auxiliary variables in the present paper. Their bias and MSE are derived. The relative efficiency of the proposed estimators as compared to relevant existing estimators of $\bar{Y}$ are examined with real data.

2. PROPOSED TRANSFORMED CHAIN RATIO TYPE ESTIMATOR OF POPULATION MEAN

A transformed chain ratio type estimator is proposed as

$$T_2 = \bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right)^{\alpha}\left(\frac{\bar{Z}}{\bar{z}'}\right)^{\beta} \tag{2.1}$$

where α and β are some unknown scalar quantities. Obviously, the estimator T_2 is biased estimator as $E(T_2) \neq \bar{Y}$. Bias of the transformed chain ratio-type estimator is as follows

$$B\left(T_2\right) = E\left(T_2\right) - \bar{Y}$$

$$= E\left[\bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right)^{\alpha}\left(\frac{\bar{Z}}{\bar{z}'}\right)^{\beta}\right] - \bar{Y}$$

We assume that

$\bar{y} = \bar{Y} + \in_0, \bar{x}' = \bar{X} + \in_1, \bar{x} = \bar{X} + \in_2, \bar{z}' = \bar{Z} + \in_3$ with $E\left(\in_0\right) = E\left(\in_1\right) = E\left(\in_2\right) = E\left(\in_3\right) = 0$

Under these assumptions and following the procedures given in Sukatme and Sukhatme (1970), an appropriate bias of T_2 upto first order of approximation is obtained as

$$B\left(T_2\right) = \bar{Y}\left[\left(\frac{1}{n} - \frac{1}{N}\right)(QC_x^2 - \alpha\rho_{yx}C_yC_x) + \left(\frac{1}{n'} - \frac{1}{N}\right)\left(PC_x^2 + RC_z^2 - \alpha^2 C_x^2 + \alpha\rho_{yx}C_yC_x - \beta\rho_{yz}C_zC_y\right)\right] \tag{2.2}$$

where

$P = \frac{\alpha\left(\alpha-1\right)}{2}$, $Q = \frac{\alpha\left(\alpha+1\right)}{2}$ and $R = \frac{\beta\left(\beta+1\right)}{2}$

If $\alpha=1$ and $\beta=1$, we find that $P=0$, $Q=1$ and $R=1$, and in this case T_2 reduces to T_1 with bias of T_1 and as given below

$$B(T_1)=\bar{Y}\left[\left(\frac{1}{n}-\frac{1}{n'}\right)\left(C_x^2-\rho_{yx}C_yC_x\right)+\left(\frac{1}{n'}-\frac{1}{N}\right)\left(C_z^2-\rho_{yz}C_zC_y\right)\right] \quad (2.3)$$

Mean Square Error (MSE) of the proposed estimator T_2

$$MSE(T_2)=E\left(T_2-\bar{Y}\right)^2$$

$$=E\left[\bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right)^{\alpha}\left(\frac{\bar{Z}}{\bar{z}'}\right)^{\beta}-\bar{Y}\right]^2$$

$$=E\left[\left(Y+\epsilon_0\right)\left(\frac{\bar{X}+\epsilon_1}{\bar{X}+\epsilon_2}\right)^{\alpha}\left(\frac{\bar{Z}}{\bar{Z}+\epsilon_3}\right)^{\beta}-\bar{Y}\right]^2$$

Under the assumptions as mentioned above, MSE of T_2 upto first order of approximation is obtained as

$$MSE(T_2)=V(\bar{y})+\bar{Y}^2\left[\alpha^2\left(\frac{1}{n}-\frac{1}{n'}\right)C_x^2+\beta^2\left(\frac{1}{n'}-\frac{1}{N}\right)C_z^2-2\alpha\left(\frac{1}{n}-\frac{1}{n'}\right)\rho_{yx}C_yC_x-2\beta\left(\frac{1}{n'}-\frac{1}{N}\right)\rho_{yz}C_yC_z\right] \quad (2.4)$$

Now, if $\alpha=1$ and $\beta=1$, we find that in this case T_2 reduces to T_1 with MSE of T_1 as given below

$$MSE(T_1)=V(\bar{y})+\bar{Y}^2\left[\left(\frac{1}{n}-\frac{1}{n'}\right)C_x^2+\left(\frac{1}{n'}-\frac{1}{N}\right)C_z^2-2\left(\frac{1}{n}-\frac{1}{n'}\right)\rho_{yx}C_yC_x-2\left(\frac{1}{n'}-\frac{1}{N}\right)\rho_{yz}C_yC_z\right] \quad (2.5)$$

We shall find out the optimum value of α and β by minimizing the $MSE(T_2)$ given in (2.4) with respect to α and β.

The above expression of $MSE(T_2)$ in (2.4) can also be written as

$$MSE(T_2)=V(\bar{y})+\bar{Y}^2\left[\alpha^2\theta_3C_x^2+\beta^2\theta_2C_z^2-2\alpha\theta_3\rho_{yx}C_yC_x-2\beta\theta_2\rho_{yz}C_yC_z\right] \quad (2.6)$$

where $\theta_1=\left(\frac{1}{n}-\frac{1}{N}\right),, \theta_2=\left(\frac{1}{n'}-\frac{1}{N}\right)\theta_3=\left(\frac{1}{n}-\frac{1}{n'}\right)$

Differentiating MSE (T_2) with respect to α, we get

$$\frac{\partial MSE(T_2)}{\partial\alpha}=\bar{Y}^2\left[2\alpha\theta_3C_X^2-2\theta_3\rho_{yx}C_yC_X\right]$$

Equating the above differential to zero and after solving it, we get the optimum value of α as

$$\alpha_{opt}=\rho_{yx}\frac{C_y}{C_x} \quad (2.7)$$

Differentiating $MSE(T_2)$ with respect to β, we get

$$\frac{\partial MSE(T_2)}{\partial\beta}=\bar{Y}^2\left[2\alpha\theta_3C_x^2-2\theta_3\rho_{yx}C_yC_x\right]$$

Equating the above differential to zero and after solving it, we get the optimum value of β as

$$\beta_{opt}=\rho_{yz}\frac{C_y}{C_z} \quad (2.8)$$

Putting the optimum value of α and β in equation (2.4), we get optimum MSE after little simplification as

$$MSE(T_2)_{opt}=V(\bar{y})-\bar{Y}^2\left[\theta_3\rho_{yx}^2C_y^2+\theta_2\rho_{yz}^2C_y^2\right] \quad (2.9)$$

3. PROPOSED COMPOSITE CHAIN RATIO TYPE ESTIMATOR OF POPULATION MEAN

We propose a composite chain ratio type estimator of population $\bar{Y}$ as

$$T_3=W\bar{y}+(1-W)T_1 \quad (3.1)$$

where W is some unknown constants.

The bias of T_3 is obtained as

$$B(T_3)=E(T_3)-\bar{Y}$$

$$=E\left[W\bar{y}+(1-W)T_1\right]-\bar{Y}$$

$$=E\left[W\bar{y}+(1-W)\bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right)\left(\frac{\bar{Z}}{\bar{z}'}\right)\right]-\bar{Y}$$

$$=E\left[W\left(\bar{Y}+\epsilon_0\right)+(1-W)\left(\bar{Y}+\epsilon_0\right)\left(\frac{\bar{X}+\epsilon_1}{\bar{X}+\epsilon_2}\right)\left(\frac{\bar{Z}}{\bar{Z}+\epsilon_3}\right)\right]-\bar{Y}$$

under the assumptions on ϵ_i's, i=1,2,3 in Section-2, the bias of T_3, upto the first order of approximation, is obtained as

$$B(T_3)=(1-W)\bar{Y}\left[\theta_3\left(C_x^2-\rho_{yx}C_yC_x\right)+\theta_2\left(C_z^2-\rho_{yz}C_yC_z\right)\right] \quad (3.2)$$

Mean square error of proposed estimator T_3 is derived as

$$MSE(T_3) = E(T_3 - \bar{Y})^2$$

$$= E\left[W\bar{y} + (1-W)T_1 - \bar{Y}\right]^2$$

$$= E\left[W\bar{y} + (1-W)\bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right)\left(\frac{\bar{Z}}{\bar{z}'}\right) - \bar{Y}\right]^2$$

$$= E\Big[W(\bar{Y} + \epsilon_0) +$$

$$(1-W)(\bar{Y} + \epsilon_0)\left(\frac{\bar{X} + \epsilon_1}{\bar{X} + \epsilon_2}\right)\left(\frac{\bar{Z}}{\bar{Z} + \epsilon_3}\right) - \bar{Y}\Big]^2$$

Upto the first order approximation, $MSE(T_3)$ is obtained as

$$MSE(T_3) = W^2\theta_1\bar{Y}^2C_y^2 + (1-W)^2\bar{Y}^2[A] + 2(1-W)\bar{Y}^2[B] \tag{3.3}$$

where

$$A = \theta_1 C_y^2 + \theta_2 C_z^2 + \theta_3 C_x^2 - 2\theta_3\rho_{yx}C_xC_y - 2\theta_2\rho_{yz}C_yC_z$$

and $B = \theta_1 C_y^2 - \theta_3\rho_{yx}C_yC_x - \theta_2\rho_{yz}C_yC_z$

Differentiating the $MSE(T_3)$ given in (3.2) with respect to W, we get

$$\frac{\partial MSE(T_3)}{\partial W} = 2W\theta_1\bar{Y}^2C_y^2 - 2(1-W)\bar{Y}^2A + 1(1-2W)\bar{Y}^2B$$

Equating the above differential to zero and after solving it, we get the optimum value of W as

$$W_{opt} = \frac{A-B}{\theta_1 C_y^2 + A - 2B} \tag{3.4}$$

Putting the value of W in equation no. (3.3), we get optimum MSE as follows

$$MSE(T_3)_{opt} = \bar{Y}^2\left[\frac{(A-B)^2\theta_1C_y^2 + (\theta_1C_y^2 - B)^2A + 2(A-B)(\theta_1C_y^2 - B)B}{(\theta_1C_y^2 + A - 2B)^2}\right] \tag{3.5}$$

4. EMPIRICAL ILLUSTRATION FOR RELATIVE EFFICIENCY OF THE PROPOSED ESTIMATORS

We have considered four real populations for the purpose of investigation of the relative efficiency of the estimators which descriptions are given below

Table 1. Description of the population

S.N.	Population size	Source of Data	Y	X	Z
1	34	Singh & Chaudhary *Theory and Analysis of Sample Survey Designs*, First Edition, 1986, pp- 177	Area Under Wheat in 1974	Area Under Wheat in 1973	Total Cultivated Area in 1971
2	34	Sukhatme & Sukhatme *Sampling theory of surveys with application*, 1970, pp -185	Area Under Wheat in 1937	Area Under Wheat in 1936	Total Cultivated Area in 1931
3	200	Sukhatme & Chand (1977)	Apple trees of bearing age in 1964	Bushels of apples harvested in 1964	Bushels of apples harvested in 1959
4	34	B.K.Singh, W. W. Chanu and Manoj Kumar (2015) *Journal of Statistics Applications & Probability* 4, No. 1, 37-51	Area under wheat in 1964	Area under wheat in 1963	Cultivated area in 1961

The details of parameters of the populations are given below

Table 2. Description of population parameters

S.N.	Parameter	Population I	Population II	Population III	Population IV
1	N	34	34	200	34
2	n′	10	10	30	10
3	n	4	4	20	7
4	C_x	0.72	0.76	2.02	.72
5	C_y	0.75	0.75	1.59	.75
6	C_z	0.85	0.62	1.44	.59
7	ρ_{yx}	0.98	0.93	0.93	.98
8	ρ_{yz}	0.44	0.89	0.77	.90
9	ρ_{zx}	0.45	0.83	0.84	.91
10	$\bar{X}$	208.88	218.41	2934.58	208.89
11	$\bar{Y}$	199.44	201.41	1031.82	199.44
12	$\bar{Z}$	856.41	765.35	3651.49	747.59

The relative efficiency of estimators is determined against the simple sample mean per unit $\bar{y}$ as

$$E_i = \frac{V(\bar{y})}{MSE(t_i)} \times 100$$

Table 3. The variance and relative efficiency (RE%) of the estimators as compared to simple sample mean $(\bar{y})$ in four different populations

Estimator	Population I		Population II		Population III		Population IV	
	Variance	RE (%)	Variance	RE(%)	Variance	RE (%)	Variance	RE (%)
$\bar{y}$	4977.47	100.00	5094.069	100.00	122168.87	100.00	2558.88	100.00
$\bar{y}_{rd}$	3446.36	144.42	3704.29	137.50	63017.20	193.80	1630.84	156.90
$\bar{y}_1$	3445.48	144.46	3692.49	137.90	81974.55	149.00	1629.91	156.99
T_1	2229.69	223.23	2411.50	211.24	60180.54	203.00	350.18	730.72
T_2	1415.42	351.66	793.64	641.86	41478.02	294.53	328.22	779.60
T_3	1578.98	315.23	884.68	575.80	40974.91	298.15	334.68	764.60
T_{l2}	3978.51	125.10	3260.18	156.66	41832.11	292.05	998.69	256.22
T_{l3}	3225.62	154.31	1848.91	275.51	49500.80	246.80	525.84	486.62
T_{l4}	3148.63	158.08	2402.59	212.59	36904.56	292.69	338.21	756.57
$\bar{Y}_{EC}^{RdR}$	2453.80	202.84	1416.61	359.59	51561.1	236.94	713.18	358.79
$\bar{Y}_{EdR}^{dc}$	4957.32	100.40	5050.92	100.85	121498.14	100.55	2588.24	98.86

where t_i's are various existing and proposed estimators. Explicitly, they are denoted as $t_1 = \bar{y}_{rd}$, $t_2 = \bar{y}_{lrd}$, $t_3 = T_1$, $t_4 = T_2$, $t_5 = T_3$, $t_6 = T_{l2}$, $t_7 = T_{l3}$, $t_8 = T_{l4}$, $t_9 = \bar{Y}_{EC}^{RdR}$ and $t_{10} = \bar{Y}_{EdR}^{dc}$

These relative efficiencies have been worked out for three populations which are described in Table 3.

It can be observed from the results of the Table 3 that the proposed estimator, i.e Transformed chain ratio type estimator has out-performed other estimators in all the populations except in population III where transformed chain ratio-type and composite chain ratio-type estimators are almost at par . For population III, transformed chain ratio-type, composite chain ratio-type, T_{l2} due to Kiregyera (1980) and T_{l4} due to Khare *et al.* (2013) are almost equally efficient.

In general, the proposed transformed chain ratio type estimator can be recommended for estimator of population mean ($\bar{Y}$) using information on two auxiliary variables. However, some more composite chain ratio cum-regression estimators can be envisaged, which is under investigation by authors.

ACKNOWLEDGEMENT

Authors are highly thankful to the referee for indicating valuable corrections which have improved the earlier version of the paper.

REFERENCES

Chand, L. (1975). "Some ratio-type estimators based on two or more auxiliary variables". Ph.D. thesis submitted to Iowa State University, Ames, IOWA.

Khare B.B., Srivastava U. and Kumar K. (2013). "A generalized chain ratio in regression estimator for population mean using two auxiliary characters in sample survey". Jour. Sci. Res., **57**, 147-153.

Kiregyera, B. (1980): "A chain ratio – type estimator in finite population double sampling using two auxiliary variables". Metrika, **27(1)**: 217-223.

Kiregyera, B. (1984). "Regression-type estimator using two auxiliary variables and model of double sampling from finite population". Metrika, **31**, 215-226.

Singh, B.K. and Choudhury, S.(2012). "Exponential chain ratio and product type estimators for finite population mean under double sampling scheme". Global Jr. of Science Frontier Research Mathematics and Decisions Sciences, 12(6).

Singh, B.K.Chanu and K. Manoj (2015). "Exponential chain ratio cum dual to Ratio estimator of finite population mean under double sampling scheme". *Jr. of Statatistics appl. & probability,* **4(1)**, 37-51.

Singh, D. & Chaudhary, F.S. (1986). Theory and Analysis of Sample Survey Designs, First Edition, pp- 177

Srivastava, S.K and Jhajj, H.S. (1995). "Classes of estimator of finite population mean and variance using auxiliary information". J. of the Indian Society of Agril. Statistics. **47(2)**:119-128.

Srivenkataramana, T. (1980). "A dual to ratio estimator in sample surveys". Biometrika, **67(1)**:199-204.

Sukhatme, B.V. and Chand L. (1977). "Multivariate ratio type estimator". Proceedings, Social Statistical Section, Amer. Stat. Association, 927-931

Sukhatme, P.V. & Sukhatme, B.V. (1970). "Sampling theory of surveys with application". Iowa State University Press, Ames, Iowa, USA.

Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 74(1) 2020 23–32

Product Type Calibration Estimators of Finite Population Total under Two Stage Sampling

Ankur Biswas, Kaustav Aditya, U.C. Sud and Pradip Basak
ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Received 19 December 2018; Revised 10 October 2019; Accepted 18 October 2019

SUMMARY

The Calibration Approach proposed by Deville and Särndal (1992) is a popular technique to efficiently use auxiliary information in survey sampling. In this study, calibration estimators of the finite population total have been developed under two stage sampling design along with variance of the estimator and the corresponding estimator of variance. It is assumed that the population level complex auxiliary information is available at the second stage of selection and the study variable is inversely related to the available auxiliary information. The proposed calibration estimators were evaluated through a simulation study and it was found that all the proposed product type calibration estimators perform better than the Horvitz-Thompson estimator as well as usual product estimator of the population total under two stage sampling design.

Keywords: Calibration approach; Auxiliary information; Product estimator; Simulation.

1. INTRODUCTION

In sample surveys, auxiliary information on the finite population is often used to increase the precision of estimators of unknown finite population parameters of study variable. In the simplest settings, ratio and regression estimators incorporate known finite population parameters of auxiliary variables in estimation of study variable parameters. The Calibration Approach, proposed by Deville and Särndal (1992), is one of the widely used techniques for incorporation of auxiliary information in estimation stages of survey sampling. In fact, the generalized regression estimator (GREG) (Cassel *et al.*, 1976) is a special case of the calibration estimator choosing the Chi-square distance function (Deville and Särndal, 1992). Calibration technique implies that a set of initial weights (usually the sampling design weights) are transformed into a set of new weights, called calibrated weights, which is the product of its initial weight and a calibration factor. In the past few decades, calibration estimation has gained significant attention not only in the field of survey methodology, but also in survey practice. Following Deville and Särndal (1992), a lot work has been carried out in calibration estimation i.e. Singh *et al.* (1998, 1999), Wu and Sitter (2001), Sitter and Wu (2002), Kott (2006) etc. Kim and Park (2010) and Särndal (2007) provided comprehensive review on calibration approach.

In many medium to large scale surveys, it is very often the case that the sampling frame is often unavailable or it could be too expensive to construct one. Also, the population could be spread over a wide area entailing very high operational expenses for personal interviews and supervisions. Two stage sampling serves as a solution in such situations where groups of elements, called primary stage units (PSU), are selected first and, then, a sample of elements, called secondary stage units (SSU), are selected from each selected PSU. For example, in agricultural surveys, villages can be selected as PSU and farmers can be selected as SSU. Estimation of the population parameters in two stage sampling using auxiliary information has been well addressed in survey sampling. Sukhatme *et al.* (1984) suggested regression estimator

Corresponding author: Ankur Biswas
E-mail address: ankur.biswas@icar.gov.in

of the population mean in two-stage sampling. Särndal *et al.* (1992) considered three different situations with respect to availability of complex auxiliary variable under two stage sampling and discussed extensively on ratio and regression estimators under such situations. Aditya *et al.* (2016a, 2016b) and Mourya *et al.* (2016) extended the calibration estimation under different cases of availability of complex auxiliary information under two stage sampling. Sinha *et al.* (2016) proposed calibration estimators for estimating population mean under stratified sampling and stratified double sampling. Aditya *et al.* (2017) attempted to use calibration approach for estimation of crop yields at the district level under two-stage sampling. Basak *et al.* (2017) proposed a calibration estimator of finite population regression coefficient under two-stage sampling design. Veronica *et al.* (2018) considered computation of calibration weights at both the first and second stages of sample selection for estimation of population mean by assuming the population means of auxiliary variables are known at both the stages of sample selection under equal probability two-stage sampling.

It was observed that most of the work related to calibration estimation for the finite population parameters were mostly restricted with the assumption of linear relationship between the study variable and the auxiliary variable. There may be situations when it can be seen that the study variable is inversely related to the auxiliary variable. For instance, an inverse relation, generally, exists between the age of individuals and hours of sleep (Sud *et al.*, 2014a). Again, in household surveys, it is often the case that marketable surplus is inversely related to family consumption for seed, feed etc. In these situations, the product estimator, proposed by Murthy (1964), is a feasible alternative. In that situation the usual methodology for calibration estimation may not fit in. Sud *et al.* (2014a, 2014b) studied the calibration approach for estimation of population total when variable of interest and auxiliary information have inverse relation under uni-stage equal probability sampling. However, multi-stage designs are most prevalent in medium to large scale surveys. Therefore, in this present study, an attempt has been made to develop calibration estimators of finite population total under two stage sampling when study variable is inversely related to the auxiliary variable.

In Section 2, proposed product type calibration estimators of finite population total under two stage sampling has been discussed. In order to study the statistical properties of proposed estimators empirically, a simulation study was carried out. Details of simulation study and discussion on simulation results are given in Section 3 and 4 respectively. Section 5 comprises concluding remarks.

2. PROPOSED CALIBRATION ESTIMATORS UNDER TWO STAGE SAMPLING DESIGN

In this section, two different calibration estimates are proposed under two stage sampling design under the assumption that available auxiliary information is inversely related to the study variable. The proposed estimators were developed with the assumption of availability of auxiliary information at SSU level under two stage sampling. Let, U be the finite population under consideration and Y be the character under study. U is grouped into N different PSUs such that $U_I = \{1,...,i,...,N\}$ and i^{th} PSU consists of M_i SSUs such that $U_i = \{1,...,k,...,M_i\}$, $i \in U_I$. Thus, we have $U = \bigcup_{i=1}^{N} U_i$ and total number of SSUs in the population U is $M_0 = \sum_{i=1}^{N} M_i$. Under two stage sampling, at stage one, a sample of PSUs, s_I, of size n PSUs is selected from U_I according to a specified design $p_I(\cdot)$ with $\pi_{Ii} = P(i \in s_I)$ and $\pi_{Iij} = P(i, j \in s_I)$ as the inclusion probabilities at the PSU level. Given that the PSU U_i is selected at the first stage, a sample s_i of size m_i SSUs is drawn from U_i according to some specified design $p_i(\cdot)$ with inclusion probabilities $\pi_{k/i} = P(k \in s_i / i \in s_I)$ and $\pi_{kl/i} = P(k, l \in s_i / i \in s_I)$ at the SSU level. In the second stage of sampling, invariance and independence property is followed. The entire sample of elements is defined as, $s = \bigcup_{i=1}^{s_I} s_i$. Let, y_{ik} denotes the observation of the study variable from k^{th} SSU in i^{th} PSU and it is observed for all $k \in s$. The parameter of interest is the population total $t_y = \sum_{i=1}^{N}\sum_{k=1}^{M_i} y_{ik} = \sum_{i=1}^{N} t_{yi}$, where $t_{yi} = \sum_{k=1}^{M_i} y_{ik} = i^{th}$ PSU total. An attempt has been made to improve the ordinary Horvitz-Thompson (1952) estimator for population total as given by

$$\hat{t}_{y\pi} = \sum_{i=1}^{n} a_{Ii} \sum_{k=1}^{m_i} \left(a_{k/i} y_{ik} \right) = \sum_{i=1}^{n} \sum_{k=1}^{m_i} a_{ik} y_{ik} \qquad (2.1)$$

where, the design weights are given as

$a_{Ii} = 1/\pi_{Ii}$, $\forall i \in s_I$, $\quad a_{k/i} = 1/\pi_{k/i}$, $\forall k \in s_i$, $i \in s_I$ and $a_{ik} = a_{Ii}.a_{k/i}$.

Under two stage sampling design, the complex auxiliary information may be available for the PSUs as well as the SSUs within the PSUs (Särndal *et al.*, 1992). In the present study, as per availability of complex auxiliary information at the ultimate stage units following two cases have been considered under two stage sampling design

Case 1: Population level complete auxiliary information is available at the SSU level.

Case 2: Population level auxiliary information is available only for the selected PSUs.

2.1 Case 1: Population level complete auxiliary information is available at SSU level

Under this case, it has been assumed that population level complete auxiliary information is available at the unit (SSU) level i.e. the auxiliary information of k^{th} SSU in i^{th} PSU, x_{ik}, is known for all elements $k \in U$. A correct value of $\sum_{i=1}^{N} \sum_{k=1}^{M_i} x_{ik}^{-1}$ is assumed to be known. In addition, there exist an inverse relationship between the study variable *Y* and the auxiliary variable *X*.

Using the well-known Calibration Approach (Deville and Särndal, 1992), we wish to modify the total design weight of k^{th} SSU of i^{th} PSU, i.e. $a_{ik} = a_{Ii}.a_{k/i}$, as given in the HT estimator of population total in Equation 2.1. The proposed product type calibration estimator of population total under **Case 1** is given by

$$\hat{t}_{yCP1} = \sum_{i=1}^{n} \sum_{k=1}^{m_i} w_{1ik} y_{ik} \qquad (2.1.1)$$

where, w_{1ik} is the calibrated weight corresponding to the design weight a_{ik} under **Case 1**.

In order to obtain the calibrated weight w_{1ik}, we minimized the Chi-square type distance $\sum_{i=1}^{n} \sum_{k=1}^{m_i} \frac{(w_{1ik} - a_{ik})^2}{a_{ik} q_{ik}}$ subject to the calibration constraint $\sum_{i=1}^{n} \sum_{k=1}^{m_i} w_{1ik} x_{ik}^{-1} = \sum_{i=1}^{N} \sum_{k=1}^{M_i} x_{ik}^{-1}$, where q_{ik} are suitably chosen constants. Using the method of Lagrange multiplier, by minimizing

$$\varphi(w_{1ik}, \lambda) = \sum_{i=1}^{n} \sum_{k=1}^{m_i} \frac{(w_{1ik} - a_{ik})^2}{a_{ik} q_{ik}} - \lambda \left(\sum_{i=1}^{n} \sum_{k=1}^{m_i} w_{1ik} x_{ik}^{-1} - \sum_{i=1}^{N} \sum_{k=1}^{M_i} x_{ik}^{-1} \right)$$

the calibrated weights are obtained as given by

$$w_{1ik} = a_{ik} + a_{ik} q_{ik} x_{ik}^{-1} \left[\frac{\sum_{i=1}^{N} \sum_{k=1}^{M_i} x_{ik}^{-1} - \sum_{i=1}^{n} \sum_{k=1}^{m_i} a_{ik} x_{ik}^{-1}}{\sum_{i=1}^{n} \sum_{k=1}^{m_i} a_{ik} q_{ik} x_{ik}^{-2}} \right],$$

$$\forall k = 1, 2, ..., m_i \text{ and } \forall i = 1, 2, .., n \qquad (2.1.2)$$

Using the results of the Equation (2.1.2) in (2.1.1) considering $q_{ik} = x_{ik}$, we have therefore proved the following result.

Theorem 1: Under **Case 1** of two stage sampling, if we consider the calibrated design weights as $w_{1ik} = a_{ik} \left(\sum_{i=1}^{N} \sum_{k=1}^{M_i} x_{ik}^{-1} \Big/ \sum_{i=1}^{n} \sum_{k=1}^{m_i} a_{ik} x_{ik}^{-1} \right)$, $\forall k = 1, 2, ..., m_i$, then the proposed product type calibration estimator of population total is given as

$$\hat{t}_{yCP1} = \sum_{i=1}^{n} \sum_{k=1}^{m_i} w_{1ik} y_{ik}$$

$$= \left(\sum_{i=1}^{n} \sum_{k=1}^{m_i} a_{ik} y_{ik} \right) \left(\sum_{i=1}^{N} \sum_{k=1}^{M_i} x_{ik}^{-1} \right) \Bigg/ \left(\sum_{i=1}^{n} \sum_{k=1}^{m_i} a_{ik} x_{ik}^{-1} \right). \qquad (2.1.3)$$

Corollary 1: Under an equal probability without replacement sampling design (Simple Random Sampling without replacement (SRSWOR)) at both the stages of two stage sampling, the proposed product type calibration estimator under **Case 1** reduces to

$$\hat{t}_{yCP1} = \left(\frac{N}{n} \sum_{i=1}^{n} \frac{M_i}{m_i} \sum_{k=1}^{m_i} y_{ik} \right) \left(\sum_{i=1}^{N} \sum_{k=1}^{M_i} x_{ik}^{-1} \right) \Bigg/ \left(\frac{N}{n} \sum_{i=1}^{n} \frac{M_i}{m_i} \sum_{k=1}^{m_i} x_{ik}^{-1} \right) \qquad (2.1.4)$$

The theoretical bias of the proposed product type calibration estimator $\hat{t}_{yCP1}$ is obtained through Taylor series linearization technique as

$$Bias\left(\hat{t}_{yCP1}\right)=\frac{\sum_{i=1}^{N}\sum_{k=1}^{M_i}y_{ik}}{\sum_{i=1}^{N}\sum_{k=1}^{M_i}x_{ik}^{-1}}\left[\frac{Cov\left(\sum_{i=1}^{n}\sum_{k=1}^{m_i}a_{ik}y_{ik},\sum_{i=1}^{n}\sum_{k=1}^{m_i}a_{ik}x_{ik}^{-1}\right)}{\sum_{k=1}^{N}\sum_{k=1}^{M_i}y_{ik}}+\frac{V\left(\sum_{i=1}^{n}\sum_{k=1}^{m_i}a_{ik}x_{ik}^{-1}\right)}{\sum_{i=1}^{N}\sum_{k=1}^{M_i}x_{ik}^{-1}}\right] \tag{2.1.5}$$

Under SRSWOR at both the stages we obtain the bias using Taylor series linearization as given by

$$Bias(\hat{t}_{yCP1})=t_y\left[\left(\frac{1}{n}-\frac{1}{N}\right)\left(\rho_b C_{by}C_{bx}+C_{bx}^2\right)+\frac{1}{n}\left(\bar{\rho}_w C_{wy}C_{wx}+C_{wx}^2\right)\right] \tag{2.1.6}$$

where,

$$\rho_b=\frac{S_{bxy}}{S_{bx}S_{by}},\quad C_{by}^2=\frac{S_{by}^2}{\bar{Y}_{N.}^2},\quad C_{bx}^2=\frac{S_{bx}^2}{\bar{X}_{N.}^2},$$

$$\bar{\rho}_w=\frac{\bar{S}_{wxy}}{\bar{S}_{wx}\bar{S}_{wy}},\quad C_{wy}^2=\frac{\bar{S}_{wy}^2}{\bar{Y}_{N.}^2},\quad C_{wx}^2=\frac{\bar{S}_{wx}^2}{\bar{X}_{N.}^2},$$

$$\bar{X}_{N.}=\frac{1}{N}\sum_{i=1}^{N}M_i\bar{X}_{i.},\quad \bar{Y}_{N.}=\frac{1}{N}\sum_{i=1}^{N}M_i\bar{Y}_{i.},$$

$$\bar{X}_{i.}=\frac{1}{M_i}\sum_{k=1}^{M_i}x_{ik},\quad \bar{Y}_{i.}=\frac{1}{M_i}\sum_{k=1}^{M_i}y_{ik},$$

$$S_{bxy}=\frac{1}{N-1}\sum_{i=1}^{N}\left(M_i\bar{Y}_{i.}-\bar{Y}_{N.}\right)\left(M_i\bar{X}_{i.}-\bar{X}_{N.}\right),$$

$$S_{by}^2=\frac{1}{N-1}\sum_{i=1}^{N}\left(M_i\bar{Y}_{i.}-\bar{Y}_{N.}\right)^2,$$

$$S_{bx}^2=\frac{1}{N-1}\sum_{i=1}^{N}\left(M_i\bar{X}_{i.}-\bar{X}_{N.}\right)^2,$$

$$\bar{S}_{wxy}=\frac{1}{N}\sum_{i=1}^{N}M_i^2\left(\frac{1}{m_i}-\frac{1}{M_i}\right)S_{ixy},$$

$$\bar{S}_{wy}^2=\frac{1}{N}\sum_{i=1}^{N}M_i^2\left(\frac{1}{m_i}-\frac{1}{M_i}\right)S_{iy}^2,$$

$$\bar{S}_{wx}^2=\frac{1}{N}\sum_{i=1}^{N}M_i^2\left(\frac{1}{m_i}-\frac{1}{M_i}\right)S_{ix}^2,$$

$$S_{ixy}=\frac{1}{M_i-1}\sum_{k=1}^{M_i}\left(x_{ik}-\bar{X}_{i.}\right)\left(y_{ik}-\bar{Y}_{i.}\right),$$

$$S_{iy}^2=\frac{1}{M_i-1}\sum_{k=1}^{M_i}\left(y_{ik}-\bar{Y}_{i.}\right)^2$$

$$S_{ix}^2=\frac{1}{M_i-1}\sum_{k=1}^{M_i}\left(x_{ik}-\bar{X}_{i.}\right)^2.$$

Usual product estimator under **Case 1** of two stage sampling considering SRSWOR at both the stages is given by

$$\hat{t}_{yP1}=\left(\frac{N}{n}\sum_{i=1}^{n}\frac{M_i}{m_i}\sum_{k=1}^{m_i}y_{ik}\right)\left(\frac{N}{n}\sum_{i=1}^{n}\frac{M_i}{m_i}\sum_{k=1}^{m_i}x_{ik}\right)\Bigg/\left(\sum_{k=1}^{N}\sum_{k=1}^{M_i}x_{ik}\right) \tag{2.1.7}$$

and, its bias is given as

$$Bias(\hat{t}_{yP1})=t_y\left[\left(\frac{1}{n}-\frac{1}{N}\right)\rho_b C_{by}C_{bx}+\frac{1}{n}\bar{\rho}_w C_{wy}C_{wx}\right] \tag{2.1.8}$$

It has been found that under SRSWOR at both the stages of a two stage sampling design under **Case 1**, product estimator $\left(\hat{t}_{yP1}\right)$ is better than usual HT estimator $\left(\hat{t}_{y\pi}\right)$ if $\rho_b\frac{C_{by}}{C_{bx}}<-\frac{1}{2}$ and $\bar{\rho}_w\frac{C_{wy}}{C_{wx}}<-\frac{1}{2}$. Under these conditions in two stage sampling design, it can be seen that

$$\left|Bias(\hat{t}_{yCP1})\right|\le\left|Bias(\hat{t}_{yP1})\right|.$$

Following Deville and Särndal (1992) and Särndal *et al.* (1992), the approximate variance of the proposed product type calibration estimator under **Case 1** by first order Taylor series linearization technique was obtained as

$$AV(\hat{t}_{yCP1})=\sum_{i=1}^{N}\sum_{j=1}^{N}\Delta_{Iij}\frac{t_{E_{i1}}}{\pi_{Ii}}\frac{t_{E_{j1}}}{\pi_{Ij}}+\sum_{i=1}^{N}\frac{1}{\pi_{Ii}}\sum_{k=1}^{M_i}\sum_{l=1}^{M_i}\Delta_{kl/i}\frac{E_{k/i}}{\pi_{k/i}}\frac{E_{l/i}}{\pi_{l/i}}, \tag{2.1.9}$$

where,

$$t_{E_{i1}}=\sum_{k=1}^{M_i}E_{k/i},\quad E_{k/i}=y_{ik}-\left(\sum_{i=1}^{N}\sum_{k=1}^{M_i}y_{ik}\Bigg/\sum_{i=1}^{N}\sum_{k=1}^{M_i}x_{ik}^{-1}\right)x_{ik}^{-1},$$

$$\Delta_{Iij}=(\pi_{Iij}-\pi_{Ii}\pi_{Ij}),\quad \Delta_{kl/i}=\pi_{kl/i}-\pi_{k/i}\pi_{l/i}.$$

Under SRSWOR design at both the stages the approximate variance reduces to

$$AV(\hat{t}_{yCP1}) = N^2\left(\frac{1}{n}-\frac{1}{N}\right)\left(S_{by}^2 + R_1^2 S_{bx^{-1}}^2 - 2R_1 S_{byx^{-1}}\right) + \frac{N}{n}\sum_{i=1}^{N} M_i^2\left(\frac{1}{m_i}-\frac{1}{M_i}\right)\left(S_{iy}^2 + R_1^2 S_{ix^{-1}}^2 - 2R_1 S_{iyx^{-1}}\right) \quad (2.1.10)$$

where,

$$R_1 = \left(\sum_{i=1}^{N}\sum_{k=1}^{M_i} y_{ik}\right)\Bigg/\left(\sum_{i=1}^{N}\sum_{k=1}^{M_i} x_{ik}^{-1}\right),$$

$$\bar{X}_{(-1)i.} = \frac{1}{M_i}\sum_{k=1}^{M_i} x_{ik}^{-1},\ \ \bar{X}_{(-1)\mathrm{N}.} = \frac{1}{N}\sum_{i=1}^{N} M_i \bar{X}_{(-1)i.},$$

$$S_{byx^{-1}} = \frac{1}{N-1}\sum_{i=1}^{N}\left(M_i\bar{Y}_{i.} - \bar{Y}_{N.}\right)\left(M_i\bar{X}_{(-1)i.} - \bar{X}_{(-1)\mathrm{N}.}\right),$$

$$S_{bx^{-1}}^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(M_i\bar{X}_{(-1)i.} - \bar{X}_{(-1)\mathrm{N}.}\right)^2,$$

$$S_{iyx^{-1}} = \frac{1}{M_i-1}\sum_{i=1}^{M_i}\left(y_{ik} - \bar{Y}_{i.}\right)\left(x_{ik}^{-1} - \bar{X}_{(-1)i.}\right),$$

$$S_{ix^{-1}}^2 = \frac{1}{M_i-1}\sum_{i=1}^{M_i}\left(x_{ik}^{-1} - \bar{X}_{(-1)i.}\right)^2.$$

Following Särndal *et al.* (1992), the Yates–Grundy form of estimator of variance (Yates and Grundy, 1953) of the proposed product type calibration estimator under **Case 1** is given by

$$\hat{V}_{YG}(\hat{t}_{yCP1}) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_{Iij}\left(\frac{\hat{t}_{E_{i1}}}{\pi_{Ii}} - \frac{\hat{t}_{E_{j1}}}{\pi_{Ij}}\right)^2 + \frac{1}{2}\sum_{j=1}^{n}\frac{1}{\pi_{Ii}}\sum_{k=1}^{m_i}\sum_{l=1}^{m_i} d_{kl/i}(w_{1ik}e_{k/i} - w_{1il}e_{l/i})^2 \quad (2.1.11)$$

where,

$$\hat{t}_{E_{i1}} = \sum_{k=1}^{m_i}\frac{e_{k/i}}{\pi_{k/i}},$$

$$e_{k\ i} = y_{ik} - \left(\sum_{i=1}^{n}\sum_{k=1}^{m_i} a_{ik}y_{ik} \Bigg/ \sum_{i=1}^{n}\sum_{k=1}^{m_i} a_{ik}x_{ik}^{-1}\right)x_{ik}^{-1},$$

$$d_{Iij} = \frac{(\pi_{Ii}\pi_{Ij} - \pi_{Iij})}{\pi_{Iij}} \text{ and } d_{kl/i} = \frac{\pi_{k/i}\pi_{l/i} - \pi_{kl/i}}{\pi_{kl/i}}.$$

Under SRSWOR design at both the stages the estimator of variance reduces to

$$\hat{V}(\hat{t}_{yCP1}) = N^2\left(\frac{1}{n}-\frac{1}{N}\right)\left(\hat{S}_{by}^2 + \hat{R}_1^2\hat{S}_{bx^{-1}}^2 - 2\hat{R}_1\hat{S}_{byx^{-1}}\right) + \frac{N}{n}\sum_{i=1}^{n} M_i^2\left(\frac{1}{m_i}-\frac{1}{M_i}\right)\left(\hat{S}_{iy}^2 + \hat{R}_1^2\hat{S}_{ix^{-1}}^2 - 2\hat{R}_1\hat{S}_{iyx^{-1}}\right) \quad (2.1.12)$$

where,

$$\hat{R}_1 = \left(\frac{N}{n}\sum_{i=1}^{n}\frac{M_i}{m_i}\sum_{k=1}^{m_i} y_{ik}\right)\Bigg/\left(\frac{N}{n}\sum_{i=1}^{n}\frac{M_i}{m_i}\sum_{k=1}^{m_i} x_{ik}^{-1}\right),$$

$$\hat{S}_{bx^{-1}}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(M_i\bar{x}_{(-1)i.} - \bar{x}_{(-1)n.}\right)^2,$$

$$\hat{S}_{by}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(M_i\bar{y}_{i.} - \bar{y}_{n.}\right)^2,$$

$$\hat{S}_{byx^{-1}} = \frac{1}{n-1}\sum_{i=1}^{n}\left(M_i\bar{y}_{i.} - \bar{y}_{n.}\right)\left(M_i\bar{x}_{(-1)i.} - \bar{x}_{(-1)n.}\right),$$

$$\hat{S}_{iy}^2 = \frac{1}{m_i-1}\sum_{i=1}^{m_i}\left(y_{ik} - \bar{y}_{i.}\right)^2,$$

$$\hat{S}_{ix^{-1}}^2 = \frac{1}{m_i-1}\sum_{i=1}^{m_i}\left(x_{ik}^{-1} - \bar{x}_{(-1)i.}\right)^2,$$

$$\hat{S}_{iyx^{-1}} = \frac{1}{m_i-1}\sum_{i=1}^{m_i}\left(y_{ik} - \bar{y}_{i.}\right)\left(x_{ik}^{-1} - \bar{x}_{(-1)i.}\right),$$

$$\bar{y}_{i.} = \frac{1}{m_i}\sum_{k=1}^{m_i} y_{ik},\ \bar{y}_{n.} = \frac{1}{n}\sum_{i=1}^{n} M_i\bar{y}_{i.},\ \bar{x}_{(-1)i.} = \frac{1}{m_i}\sum_{k=1}^{m_i} x_{ik}^{-1},$$

$$\bar{x}_{(-1)n.} = \frac{1}{n}\sum_{i=1}^{n} M_i\bar{x}_{(-1)i}.$$

2.2 Case 2: Population level auxiliary information is available only for the selected PSUs

In this case, it has been assumed that the population level auxiliary information is available at the SSU level only for the selected PSUs i.e. the auxiliary information is known for all the SSUs within the PSU $i \in s_I$. The correct value of $\sum_{k=1}^{M_i} x_{ik}^{-1}$ is assumed to be available for each i^{th} sampled PSU. Suppose, there exist inverse relationship between the study variable Y and the auxiliary variable X. Using well-known Calibration Approach (Deville and Särndal, 1992), the design weight at the second stage $a_{k/i}$ has been revised. The proposed product type calibration estimator of population total under **Case 2** is given by

$$\hat{t}_{yCP2} = \sum_{i=1}^{n} a_{Ii} \sum_{k=1}^{m_i} w_{2ik} y_{ik} \tag{2.2.1}$$

where, w_{2ik} is the calibrated weight corresponding to the design weight $a_{k/i}$.

In this situation, we minimized the Chi-square type distance function $\sum_{k=1}^{m_i} \frac{(w_{2ik} - a_{k/i})^2}{a_{k/i}\, q_{ik}}$ subject to $\sum_{k=1}^{m_i} w_{2ik}\, x_{ik}^{-1} = \sum_{k=1}^{M_i} x_{ik}^{-1}$, where q_{ik} are suitably chosen constants. Using Lagrange multiplier technique, by minimizing

$$\varphi(w_{2ik}, \lambda) = \sum_{k=1}^{m_i} \frac{(w_{2ik} - a_{k/i})^2}{a_{k/i}\, q_{ik}} - \lambda \left(\sum_{k=1}^{m_i} w_{2ik}\, x_{ik}^{-1} - \sum_{k=1}^{M_i} x_{ik}^{-1} \right),$$

the new set of calibrated weights is obtained as

$$w_{2ik} = a_{k/i} + a_{k/i}\, q_{ik} x_{ik}^{-1} \left[\frac{\sum_{k=1}^{M_i} x_{ik}^{-1} - \sum_{k=1}^{m_i} a_{k/i}\, x_{ik}^{-1}}{\sum_{k=1}^{m_i} a_{k/i}\, q_{ik} x_{ik}^{-2}} \right]$$

$$\forall\, k = 1, 2, \ldots, m_i. \tag{2.2.2}$$

Using the results of the Equation (2.2.2) in (2.2.1) considering $q_{ik} = x_{ik}$, we have therefore proved the following result.

Theorem 2: Under **Case 2** of two stage sampling, if we consider the calibrated design weights as $w_{2ik} = a_{k/i} \left(\sum_{k=1}^{M_i} x_{ik}^{-1} \Big/ \sum_{k=1}^{m_i} a_{k/i}\, x_{ik}^{-1} \right), \quad \forall\, k = 1, 2, \ldots, m_i$

then the proposed product type calibration estimator of population total is given as

$$\hat{t}_{yCP2} = \sum_{i=1}^{n} a_{Ii} \sum_{k=1}^{m_i} w_{2ik} y_{ik}$$

$$= \sum_{i=1}^{n} a_{Ii} \frac{\left(\sum_{k=1}^{m_i} a_{k/i} y_{ik} \right) \left(\sum_{i=1}^{M_i} x_{ik}^{-1} \right)}{\left(\sum_{i=1}^{m_i} a_{k/i} x_{ik}^{-1} \right)}. \tag{2.2.3}$$

Corollary 2: Under SRSWOR at both the stages of two stage sampling, the proposed product type calibration estimator under **Case 2** reduces to

$$\hat{t}_{yCP2} = \frac{N}{n} \sum_{i=1}^{n} \left[\left(\frac{M_i}{m_i} \sum_{k=1}^{m_i} y_{ik} \right) \left(\sum_{i=1}^{M_i} x_{ik}^{-1} \right) \Big/ \left(\frac{M_i}{m_i} \sum_{k=1}^{m_i} x_{ik}^{-1} \right) \right]. \tag{2.2.4}$$

Using Taylor series linearization technique, its bias is obtained as

$$Bias(\hat{t}_{yCP2}) = t_y \left[\frac{1}{n} \left(\bar{\rho}_w C_{wy} C_{wx} + C_{wx}^2 \right) \right] \tag{2.2.5}$$

where, the terms are as defined in **Case 1** (Eqn. 2.1.6).

Usual product estimator under **Case 2** of two stage sampling with SRSWOR at both the stages is given by

$$\hat{t}_{yP2} = \left(\frac{N}{n} \sum_{i=1}^{n} \frac{M_i}{m_i} \sum_{k=1}^{m_i} y_{ik} \right) \left(\frac{N}{n} \sum_{i=1}^{n} \frac{M_i}{m_i} \sum_{k=1}^{m_i} x_{ik} \right) \Big/ \left(\frac{N}{n} \sum_{i=1}^{n} \sum_{k=1}^{M_i} x_{ik} \right) \tag{2.2.6}$$

and, using Taylor series linearization technique, its bias is given by

$$Bias(\hat{t}_{yP2}) = t_y \left[\frac{1}{n} \bar{\rho}_w C_{wy} C_{wx} \right] \tag{2.2.7}$$

It has been found that under SRSWOR at both the stages of two stage sampling design under **Case 2**, product estimator $\left(\hat{t}_{yP2}\right)$ is better than usual HT estimator $\left(\hat{t}_{y\pi}\right)$ if $\bar{\rho}_w \frac{C_{wy}}{C_{wx}} < -\frac{1}{2}$. Under this condition in two stage sampling design, it can be seen that

$$\left| Bias(\hat{t}_{yCP2}) \right| \le \left| Bias(\hat{t}_{yP2}) \right|.$$

Following Särndal *et al.* (1992) the approximate variance of the proposed product type calibration estimator under **Case 2** by first order Taylor series linearization technique was obtained as

$$AV(\hat{t}_{yCP2}) = \sum_{i=1}^{N}\sum_{j=1}^{N} \Delta_{Iij} \frac{t_{E_{i2}}}{\pi_{Ii}} \frac{t_{E_{j2}}}{\pi_{Ij}} + \sum_{i=1}^{N} \frac{1}{\pi_{Ii}} \sum_{k=1}^{M_i}\sum_{l=1}^{M_i} \Delta_{kl/i} \frac{E_{k/i}}{\pi_{k/i}} \frac{E_{l/i}}{\pi_{l/i}}, \quad (2.2.8)$$

where,

$$t_{E_{i2}} = \sum_{k=1}^{M_i} E_{k/i}\ ,\ E_{k/i} = y_{ik} - R_i x_{ik}^{-1},\ .$$

$$R_i = \left(\sum_{i=1}^{M_i} y_{ik} \Bigg/ \sum_{i=1}^{M_i} x_{ik}^{-1}\right).$$

Under SRSWOR design at both the stages, it reduces to

$$AV(\hat{t}_{yCP2}) = N^2\left(\frac{1}{n} - \frac{1}{N}\right) S_{by}^2 + \frac{N}{n}\sum_{i=1}^{N} M_i^2 \left(\frac{1}{m_i} - \frac{1}{M_i}\right)\left\{S_{iy}^2 + R_i^2 S_{ix^{-1}}^2 - 2R_i S_{iyx^{-1}}\right\} \quad (2.2.9)$$

where, the terms are as defined in **Case 1** (Eqn. 2.1.6 and 2.1.10).

The Yates–Grundy form of estimator of variance of the proposed product type calibration estimator under **Case 2** is given by

$$\hat{V}_{YG}(\hat{t}_{yCP2}) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_{Iij}\left(\frac{\hat{t}_{E_{i2}}}{\pi_{Ii}} - \frac{\hat{t}_{E_{j2}}}{\pi_{Ij}}\right)^2 + \frac{1}{2}\sum_{j=1}^{n} \frac{1}{\pi_{Ii}} \sum_{k=1}^{m_i}\sum_{l=1}^{m_i} d_{kl/i}(w_{2ik}e_{k/i} - w_{2il}e_{l/i})^2 \quad (2.2.10)$$

where,

$$\hat{t}_{E_{i2}} = \sum_{k=1}^{n_i} \frac{e_{k/i}}{\pi_{k/i}},\ ,\ e_{k/i} = y_{ik} - \hat{R}_i x_{ik}^{-1},$$

$$\hat{R}_i = \left(\sum_{k=1}^{n_i} a_{k/i} y_{ik}\right) \Bigg/ \left(\sum_{i=1}^{n_i} a_{k/i} x_{ik}^{-1}\right),$$

$d_{Iij} = (\pi_{Ii}\pi_{Ij} - \pi_{Iij})/\pi_{Iij}$ and

$d_{kl/i} = (\pi_{k/i}\pi_{l/i} - \pi_{kl/i})/\pi_{kl/i}$.

Under SRSWOR design at both the stages it reduces to

$$\hat{V}(\hat{t}_{yCP2}) = N^2\left(\frac{1}{n} - \frac{1}{N}\right)\hat{S}_{by}^2 + \frac{N}{n}\sum_{i=1}^{n} M_i^2\left(\frac{1}{m_i} - \frac{1}{M_i}\right)\left\{\hat{S}_{iy}^2 + \hat{R}_i^2 \hat{S}_{ix^{-1}}^2 - 2\hat{R}_i \hat{S}_{iyx^{-1}}\right\} \quad (2.2.11)$$

where, the terms are as defined in **Case 1** (Eqn. 2.1.12).

3. SIMULATION STUDY

In order to evaluate the statistical performance of proposed product type calibration estimators, a simulation study was carried out. We have considered the case of two stage sampling where sample selection at each stage is governed by SRSWOR for the situation that the size of the PSU and the corresponding SSUs were fixed. For the simulation study, a finite population of 5000 units considering, number of PSU, N=50 and PSU size, M_i=100, was generated from $y_k = \beta x_k^{-1} + e_k$, $k = 1, ..., M_0$, where $M_0 = \sum_{i=1}^{N} M_i$. The auxiliary variable was generated from normal distribution with mean 5 and variance 1 i.e. $x_k \sim N(5, 1)$ and the errors, e_k, $k = 1,...,M_0$, from normal distribution with mean 0 and variance $\sigma^2 x_k^{-1}$ i.e. $e_k \sim N(0, \sigma^2 x_k^{-1})$. We have fixed the value of β = 20 and chosen four different values for σ^2 as 0.25, 1.0, 2.0 and 5.0. In this way, we generated four sets of population, denoted as **Set 1**, **Set 2**, **Set 3** and **Set 4**, with different correlation coefficient values between study variable Y and auxiliary variable X as -0.91, -0.85, -0.78 and -0.64 respectively. The value of left hand side of the Condition 1 and Condition 2 i.e. $\rho_A \frac{C_{by}}{C_{bx}} < -\frac{1}{2}$ and $\bar{\rho}_w \frac{C_{wy}}{C_{wx}} < -\frac{1}{2}$ are lesser than -0.5 in all the population sets which can be seen in the following table:

Set	Set 1	Set 2	Set 3	Set 4
Condition 1	-1.1	-1.17	-1.23	-1.34
Condition 2	-1.11	-1.1	-1.1	-1.08

Then, from each of the study population sets, we have selected a total of 10000 different samples of following sizes using SRSWOR at both the stages of the two stage sampling design and calculated different estimates of population total under **Case 1** and **2**:

n =10, m_i =20	n =15, m_i =25	n =20, m_i=30	n =25, m_i=40
n =10, m_i =25	n =15, m_i =30	n =20, m_i=40	n =25, m_i=50

Developed product type calibration estimators as well as all other usual estimators of population total under two stage sampling were evaluated on the basis of two measures viz. percentage Relative Bias (%RB) and percentage Relative Root Mean Squared Error (%RRMSE) of any estimator of the population parameter θ as given by

$$RB(\hat{\theta}) = \frac{1}{S}\sum_{i=1}^{S}\left(\frac{\hat{\theta}_i - \theta}{\theta}\right) \times 100 \text{ and}$$

$$RRMSE(\hat{\theta}) = \sqrt{\frac{1}{S}\sum_{i=1}^{S}\left(\frac{\hat{\theta}_i - \theta}{\theta}\right)^2} \times 100$$

where, $\hat{\theta}_i$ are the estimates of population parameter θ for the character under study obtained at i^{th} sample in the simulation study.

4. RESULTS AND DISCUSSION

Table 1 shows the %RB of the HT estimators ($\hat{t}_{y\pi}$), product estimators ($\hat{t}_{yP1}$ and $\hat{t}_{yP2}$), ratio estimators ($\hat{t}_{yR1}$ and $\hat{t}_{yR2}$), linear regression estimators ($\hat{t}_{ylr1}$ and $\hat{t}_{ylr2}$) (as in Särndal *et al.*, (1992), pp-323) and proposed product type calibration estimators ($\hat{t}_{yCP1}$ and $\hat{t}_{yCP2}$) of population total under both the **Case 1** and **2** respectively when available auxiliary variable is inversely related with the study variable. Table 2 presents comparison of performance of all the estimators for all the population Sets on the basis of %RRMSE.

From Table 1 it can be seen that, the proposed product type calibration estimators of the population total for both the **Case 1** and **2** of availability of auxiliary information were giving consistently least amount %RB in all the sets compared to their usual linear regression, product, ratio and HT estimators under two stage sampling design when available auxiliary variable is inversely related with the study variable. It is evident that ratio estimator is not at all suitable for this situation.

A close look of Table 2 reveals that, the product type calibration estimators of the population total developed under two stage sampling design under **Case 1** and **2** were always more efficient than the respective linear regression, product, ratio and HT estimators in all the population sets with respect to %RRMSE. The %RRMSE of both the proposed product type calibration estimators of the population total under **Case 1** and **2** were decreasing with the increase of sample sizes. With the increase of negative correlation between the study and auxiliary variable, %RRMSE of both the proposed product type calibration estimators of the population total under **Case 1** and **2** were decreasing. The proposed product type calibration estimators of the population total developed under **Case 1** of two stage sampling design was producing least %RRMSE in all Sets. Therefore, for the situations of availability of population level complete auxiliary information at SSU level i.e. **Case 1**, performance of the proposed product type calibration estimator is best among all other competitors. On the other hand, for more practical situation of availability of population level auxiliary information only for selected PSUs i.e. **Case 2**, proposed product type calibration estimator can be preferred over usual HT, product and linear regression estimators of population total.

5. CONCLUSIONS

In this study, following the calibration approach (Deville and Särndal, 1992), we proposed product type calibration estimators of the finite population total under two stage sampling design when the available auxiliary variable is inversely related to the study variable. Here, two different cases under two stage sampling viz. "***Case 1**: population level complete auxiliary information is available at the SSU level*" and "***Case 2**: population level auxiliary information is available only for the selected PSUs*" have been considered. In order to study the statistical performance of proposed product type calibration estimators as compared to existing estimators of population total of study variable, a simulation study was carried out. The simulation results show that the proposed product type calibration estimator of the population total were performing better than usual linear regression, product and HT estimators under two stage sampling design when available auxiliary variable is inversely related with the study variable. The proposed product type calibration estimators of the population total developed under **Case 1** performs better than that of **Case 2**, since more auxiliary information was available under **Case 1**.

ACKNOWLEDGEMENT

Authors are profoundly thankful to the reviewers for their productive comments for improvements.

REFERENCES

Aditya, K., Sud, U.C., Chandra, H. and Biswas, A. (2016a). Calibration Based Regression Type Estimator of the Population Total under Two Stage Sampling Design. *Journal of Indian Society of Agriculture Statistics*, **70 (1)**, 19-24.

Aditya, K., Sud, U.C. and Chandra, H. (2016b). Calibration Approach based Estimation of Finite Population Total under Two Stage Sampling. *J. Ind. Soc. Agril. Statist.*, **70(3)**, 219-226.

Aditya K., Biswas A., Gupta, A.K. and Chandra, H. (2017). District-level crop yield estimation using calibration approach. *Current Science*, **112(9)**, 1927-1931.

Basak P., Sud, U.C. and Chandra, H. (2017). Calibration Estimation of Regression Coefficient for Two-stage Sampling Design. *J. Ind. Soc. Agril. Statist.*, **71(1)**, 1-6.

Cassel, C.M., Sarndal, C.E. and Wretman, J.H. (1976). Some results on generalized difference estimation and generalized regression estimation for finite population. *Biometrika*, **63**, 615-620.

Deville, J.C. and Sarndal, C.E. (1992). Calibration estimators in survey sampling. *J. Amer. Statist. Assoc.*, **87**, 376-382.

Horvitz, D.G. and Thompson, D.J. (1952). A Generalization of Sampling without Replacement from a Finite Universe. *J. Amer. Statist. Assoc.*, **47,** 663-685.

Kim, J.K. and Park, M. (2010). Calibration estimation in survey sampling. *International Statistical Review*, **78**, 21-39.

Kott, P.S. (2006). Using calibration weighting to adjust for nonresponse and coverage errors. *Survey Methodology*, **32**, 133142.

Mourya, K.K., Sisodia, B.V.S. and Chandra, H. (2016). Calibration approach for estimating finite population parameter in two stage sampling. *Journal of Statistical Theory and Practice*, **10(3)**, 550-562.

Murthy, M.N. (1964). Product method of estimation. *Sankhya*, **26A**, 69-74.

Särndal, C.E., Swensson, B. and Wretman, J. (1992). *Model-Assisted Survey Sampling*. Springer-Verlag.

Särndal, C.E. (2007). The calibration approach in survey theory and practice. *Survey Methodology*, **33**, 99-119.

Singh, S., Horn, S. and Yu, F. (1998). Estimation of variance of the general regression estimator: Higher level calibration approach. *Survey Methodology*, **24**, 41-50.

Singh, S., Horn, S., Choudhury, S. and Yu, F. (1999). Calibration of the estimators of variance, *Australian and New Zealand Journal of Statistics,* **41(2)**, 199-212.

Sinha N., Sisodia, B.V.S., Singh, S. and Singh, S.K. (2016). Calibration Approach Estimation of Mean in Stratified Sampling and Stratified Double Sampling, *Communications in Statistics - Theory and Methods*, **46(10)**, DOI: 10.1080/03610926.2015.1091083.

Sitter, R.R. and Wu, C. (2002). Efficient estimation of quadratic finite population functions. *J. Amer. Statist. Assoc.*, **97**, 535-543.

Sud, U.C., Chandra, H. and Gupta, V.K. (2014a). Calibration based product estimator in single and two phase sampling. *Journal of Statistical Theory and Practice*, **8(1)**, 1-11.

Sud, U.C., Chandra, H. and Gupta, V.K. (2014b). Calibration approach based regression type estimator for inverse relationship between study and auxiliary variable. *Journal of Statistical Theory and Practice*, **8(4)**, 707-721.

Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S. and Asok. C. (1984). *Sampling Theory with Applications*. Indian Society of Agricultural Statistics, New Delhi.

Veronica I.S., Stephen A.S. and Sarjinder Singh. (2018). Calibrated estimators in two-stage sampling. *Communications in Statistics - Theory and Methods*, DOI: 10.1080/03610926.2018.1433850.

Wu, C. and Sitter, R.R. (2001). A model calibration approach to using complete auxiliary information from survey data. *J. Amer. Statist. Assoc.*, **96**, 185-193.

Yates, F. and Grundy, P.M. (1953). Selection without replacement from within strata with probability proportional to size. *J. Roy. Statist. Soc., **B** 9*, 223-261.

Table 1. Comparison of all the estimators under Case 1 and 2 with respect to %RB in case of all four population sets when available auxiliary variable is inversely related with the study variable

Set	Sample Size (n_m_i)	$\hat{t}_{y\pi}$	Case 1				Case 2			
			$\hat{t}_{yCP1}$	$\hat{t}_{yP1}$	$\hat{t}_{ylr1}$	$\hat{t}_{yR1}$	$\hat{t}_{yCP2}$	$\hat{t}_{yP2}$	$\hat{t}_{ylr2}$	$\hat{t}_{yR2}$
Set 1 (ρ= -0.91)	10_20	-0.016	0.001	-0.026	-0.057	0.013	-0.005	-0.029	-0.058	0.014
	10_25	0.015	-0.001	-0.011	-0.035	0.055	0.000	-0.009	-0.031	0.050
	15_25	0.000	0.002	-0.009	-0.024	0.019	0.000	-0.006	-0.020	0.015
	15_30	-0.005	0.002	-0.015	-0.026	0.012	0.002	-0.011	-0.021	0.007
	20_30	0.013	-0.001	-0.005	-0.016	0.037	0.001	-0.003	-0.013	0.034
	20_40	0.010	0.003	0.000	-0.007	0.024	0.006	0.003	-0.003	0.020
	25_40	-0.001	-0.001	-0.008	-0.013	0.009	-0.009	-0.013	-0.018	0.013
	25_50	-0.007	-0.002	-0.005	-0.008	-0.007	-0.007	-0.009	-0.012	-0.004
Set 2 (ρ= -0.85)	10_20	-0.004	0.001	-0.024	-0.054	0.034	0.006	-0.016	-0.044	0.024
	10_25	0.006	0.000	-0.018	-0.041	0.044	-0.006	-0.020	-0.042	0.042
	15_25	-0.026	-0.011	-0.025	-0.039	-0.018	-0.011	-0.026	-0.039	-0.018
	15_30	-0.015	-0.004	-0.020	-0.030	-0.002	-0.009	-0.021	-0.032	-0.002
	20_30	-0.024	-0.006	-0.019	-0.026	-0.024	-0.004	-0.017	-0.024	-0.026
	20_40	0.003	0.001	-0.008	-0.014	0.019	-0.003	-0.009	-0.015	0.019
	25_40	0.024	0.003	0.001	-0.006	0.050	0.003	0.002	-0.005	0.048
	25_50	-0.010	-0.001	-0.004	-0.007	-0.012	-0.002	-0.006	-0.008	-0.012

Set 3 (ρ= -0.78)	10_20	-0.014	-0.005	-0.022	-0.047	0.012	-0.011	-0.023	-0.047	0.011
	10_25	-0.008	0.000	-0.023	-0.044	0.021	0.000	-0.019	-0.039	0.015
	15_25	0.016	0.000	-0.010	-0.027	0.052	0.001	-0.007	-0.022	0.047
	15_30	-0.001	0.006	-0.003	-0.013	0.009	0.006	0.001	-0.009	0.004
	20_30	-0.011	-0.015	-0.019	-0.029	0.002	-0.014	-0.020	-0.029	0.002
	20_40	-0.008	0.004	-0.004	-0.008	-0.009	0.000	-0.007	-0.011	-0.007
	25_40	0.004	0.004	0.001	-0.003	0.009	0.002	0.000	-0.004	0.010
	25_50	-0.001	-0.005	-0.008	-0.011	0.007	-0.002	-0.005	-0.007	0.004
Set 4 (ρ= -0.64)	10_20	0.005	0.002	-0.021	-0.049	0.048	0.007	-0.017	-0.043	0.042
	10_25	0.031	0.030	0.016	-0.003	0.061	0.031	0.020	0.003	0.054
	15_25	-0.014	-0.014	-0.023	-0.036	0.005	-0.018	-0.023	-0.036	0.004
	15_30	0.018	0.016	0.003	-0.008	0.041	0.012	0.004	-0.007	0.039
	20_30	-0.007	-0.014	-0.017	-0.025	0.009	-0.019	-0.023	-0.031	0.014
	20_40	-0.008	-0.007	-0.013	-0.018	0.001	-0.009	-0.012	-0.017	0.000
	25_40	0.006	-0.003	-0.007	-0.012	0.022	-0.002	-0.005	-0.009	0.018
	25_50	-0.008	-0.010	-0.011	-0.014	-0.003	-0.011	-0.011	-0.014	-0.003

Table 2. Comparison of all the estimators under Case 1 and 2 with respect to %RRMSE in case of all four population sets of all the estimators when available auxiliary variable is inversely related with the study variable

Set	Sample Size (n_m_i)	$\hat{t}_{y\pi}$	Case 1				Case 2			
			$\hat{t}_{yCP1}$	$\hat{t}_{yP1}$	$\hat{t}_{ylr1}$	$\hat{t}_{yR1}$	$\hat{t}_{yCP2}$	$\hat{t}_{yP2}$	$\hat{t}_{ylr2}$	$\hat{t}_{yR2}$
Set 1 (ρ= -0.91)	10_20	1.677	0.377	0.726	0.714	2.975	0.690	0.899	0.890	2.841
	10_25	1.469	0.333	0.638	0.630	2.613	0.677	0.827	0.819	2.461
	15_25	1.199	0.271	0.524	0.513	2.112	0.521	0.659	0.651	2.011
	15_30	1.059	0.246	0.466	0.457	1.873	0.516	0.619	0.613	1.750
	20_30	0.921	0.212	0.404	0.395	1.626	0.420	0.522	0.516	1.535
	20_40	0.770	0.176	0.343	0.336	1.355	0.401	0.473	0.469	1.246
	25_40	0.663	0.154	0.299	0.294	1.170	0.332	0.399	0.397	1.091
	25_50	0.576	0.133	0.258	0.253	1.013	0.322	0.369	0.365	0.917
Set 2 (ρ= -0.85)	10_20	1.781	0.755	0.967	0.958	3.002	0.971	1.116	1.111	2.876
	10_25	1.562	0.668	0.848	0.841	2.639	0.892	1.011	1.006	2.486
	15_25	1.260	0.536	0.682	0.676	2.128	0.708	0.804	0.801	2.018
	15_30	1.139	0.485	0.615	0.610	1.931	0.666	0.746	0.744	1.812
	20_30	0.983	0.420	0.537	0.532	1.656	0.556	0.635	0.633	1.567
	20_40	0.829	0.356	0.454	0.448	1.392	0.518	0.574	0.572	1.279
	25_40	0.716	0.307	0.389	0.385	1.209	0.429	0.478	0.476	1.128
	25_50	0.620	0.267	0.337	0.333	1.045	0.398	0.434	0.432	0.949
Set 3 (ρ= -0.78)	10_20	1.907	1.081	1.220	1.215	3.048	1.229	1.343	1.341	2.924
	10_25	1.708	0.954	1.076	1.071	2.744	1.131	1.218	1.217	2.601
	15_25	1.372	0.763	0.864	0.860	2.209	0.888	0.964	0.961	2.102
	15_30	1.246	0.700	0.794	0.789	1.990	0.840	0.909	0.908	1.869
	20_30	1.073	0.588	0.667	0.662	1.731	0.696	0.751	0.749	1.644
	20_40	0.894	0.497	0.561	0.557	1.439	0.621	0.664	0.663	1.329
	25_40	0.787	0.435	0.493	0.489	1.263	0.532	0.572	0.570	1.179
	25_50	0.672	0.380	0.427	0.424	1.073	0.484	0.513	0.512	0.979
Set 4 (ρ= -0.64)	10_20	2.316	1.685	1.767	1.767	3.353	1.807	1.876	1.877	3.222
	10_25	2.063	1.510	1.576	1.574	2.982	1.635	1.690	1.690	2.840
	15_25	1.665	1.220	1.286	1.284	2.394	1.314	1.366	1.366	2.288
	15_30	1.505	1.090	1.143	1.139	2.177	1.192	1.239	1.237	2.055
	20_30	1.271	0.931	0.975	0.972	1.837	1.007	1.048	1.047	1.749
	20_40	1.072	0.777	0.810	0.807	1.553	0.869	0.895	0.895	1.445
	25_40	0.947	0.697	0.725	0.722	1.364	0.767	0.789	0.788	1.283
	25_50	0.822	0.603	0.626	0.623	1.180	0.679	0.699	0.698	1.084

Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 74(1) 2020 33–40

Bayesian State-space Implementation of Schaefer Production Model for Assessment of Stock Status for Multi-gear Fishery

Eldho Varghese, T. V. Sathianandan, J. Jayasankar, Somy Kuriakose, K.G. Mini and M. Muktha

ICAR-Central Marine Fisheries Research Institute, Kochi

Received 30 March 2019; Revised 13 August 2019; Accepted 18 October 2019

SUMMARY

Knowing the status of marine fish stock is of utmost importance to develop management strategies for sustainable harvest of marine fishery resources. A widely accepted approach towards this is to derive sustainable harvest levels using time series data on fish catch and fishing effort based on fish stock assessment models like Schaefer's model that describe the biomass dynamics. In India, the marine fishery is of complex multi-species nature where in different species are caught by a number of fishing gears and each gear harvests a number of species making it difficult to obtain the fishing effort corresponding to each fish species. Since the capacity of the gears varies, the effort made to catch a resource cannot be considered as the sum of efforts expended by different fishing gears. Hence, it demands the importance of effort standardisation for making use in stock assessment models. This paper describes a methodology for the standardization of fishing efforts and assessing fish stock status using Bayesian state-space implementation of the Schaefer production model (BSM). A Monte Carlo based method namely Catch-Maximum Sustainable Yield (CMSY), has also been used for estimating fisheries reference points from landings and a proxy for biomass using resilience of the species. The procedure has been illustrated with data on Indian mackerel (*Rastrelliger Kanagurta*) collected from the coastal state of Andhra Pradesh, India during 1997-2018. Maximum Sustainable Yield (MSY) of Indian mackerel for Andhra Pradesh has been estimated. A comparison between both CMSY and BSM methods have been made and found that the estimates are in close agreements.

Keywords: capture fisheries, stock assessment, Schaefer model, Monte Carlo method, CMSY, BSM.

1. INTRODUCTION

Knowing the status of marine stock is of utmost importance to develop management strategies for sustainable harvest of marine resources. Stock assessment is the process of collecting, analysing and reporting fish population information to determine changes in the abundance of fishery stocks in response to fishing and, to the extent possible, predict future trends of stock abundance (Sparre and Venema, 1992). In fisheries where there are no fishery-independent measures of abundance, the commercial catch rate is commonly used as an abundance indicator (Vivekanandan, 2005).

Surplus production models, introduced by (Graham, 1935) are commonly used for assessing the state of fish stocks. These models view population as one unit of biomass, with all individuals having the same growth and mortality rates. The surplus production models deal with the entire stock, the entire fishing effort and the total yield obtained from the stock. It is used to determine the optimum level of effort that is the effort that produces the maximum yield that can be sustained without affecting the long-term productivity of the stock, or the maximum sustainable yield (MSY).

Surplus production models assume that variation in population biomass results from increases due to growth and reproduction, and decreases from natural and fishing mortality. Surplus production models use Catch-Per-Unit-Effort (CPUE) as input. The data, which represent a time series of years, are usually collected from commercial fishery. The model is based on the assumption that the CPUE is proportional to biomass of the fish in the sea.

Schaefer model is one of the most popular surplus production model which gives by following equation:

Corresponding author: Eldho Varghese
E-mail address: eldhoiasri@gmail.com

$$B_{t+1} = B_t + rB_t(1 - \frac{B_t}{k}) - C_t,\ \ C_t = qE_tB_t$$

where B_{t+1} is the exploited biomass in the subsequent year t+1, B_t is the current biomass, r is the intrinsic growth rate, k is the carrying capacity, C_t is the catch in the current year t, E_t is the fishing effort at time t and q is the catchability coefficient. Surplus production models use CPUE as an index of biomass (i.e., $CPUE_t = qB_t$).

The above equation has been modified to account for reduced recruitment at severely depleted stock sizes, a linear decline of surplus production, which is a function of recruitment, somatic growth and natural mortality is incorporated if biomass falls below ¼ k (Froese *et al.*, 2017).

$$B_{t+1} = B_t + 4\frac{B_t}{k}(1 - \frac{B_t}{k})rB_t - C_t,\ \ if\ \frac{B_t}{k} < 0.25$$

The term 4*Bt/k* assumes a linear decline of recruitment below half of the biomass that is capable of producing MSY.

A major challenge in fitting such a production model is to find out CPUE, may be in terms of the units operated or in terms of hours of operation/actual fishing hours (AFH). As the fishing fleet is heterogeneous in most of the cases, it is partitioned into boat-gear categories in each of which the fishing units have similar characteristics and performance. When it comes to measure the combined effect of the fishing operations of the entire fleet to the exploitation of a fish stock, it becomes apparent that adding together effort exerted by different boat-gear categories is not always meaningful without first applying effort adjustment to increase its comparability (Stamatopoulos and Abdallah, 2015).

Stock assessment of individual species becomes difficult when a species is targeted by various gears and each gear may harvest more than the species targeted. Since the capacity of the gears vary and also each gear may contain multiple species, the effort made to catch a resource cannot be considered as the sum of duration/units of operation of all the gears. Hence, the problem of exploitation of the same stock by gears with different efficiencies has to be addressed. There are several techniques for dealing with such situations, the most commonly used one is the standardization of fishing effort. There is a lot of literature available on the standardization of the fishing effort. These methods deeply depend on characteristics of the gear being operated and the availability of the information.

Hilborn and Walters (1992) proposed the use of Generalized Linear Models (GLM) for standardization of fishing effort. Rochman *et al.* (2017) attempted to standardize CPUE to estimate relative abundance indices based on the Indonesian longline dataset time series using GLM with Tweedie distribution. Daniel *et al.* (2016) gave a method named multi-gear mean standardization (MGMS) which combining catch per unit effort data that standardizes catch per unit effort data across gear types. Setyadji *et al.* (2018) used GLM to standardize CPUE and to estimate relative abundance indices based on the Indonesian longline dataset. Six GLM models were considered viz., negative binomial, zero inflated Poisson, zero-inflated negative binomial, Poisson hurdle, and negative binomial hurdle models. AIC and BIC were used to select the best models among all those evaluated.

In the literature cited above, either CPUE or effort exerted to a catch particular fish or vessel/ gear characteristics available. A methodology for the standardization of fishing effort is to be required when one has to estimate the effort exerted to catch a particular species from the total effort hence it demands the importance of effort standardisation for making use in stock assessment models. Here, an attempt has been made to develop a methodology to standardize the fishing effort and further to arrive at MSY using Bayesian approach. A Monte Carlo method has also been used to obtain the MSY when a measure of fishing effort is not available. This is done by making use of the species resilience to derive a quantitative measure of productivity.

2. COMPUTATIONAL STEPS FOR EFFORT STANDARDIZATION

This method of standardization requires the species catch, total catch and total fishing effort. Let Y_{ijk} represents the catch of k^{th} species (k=1,2,...,s) from i^{th} (i=1,2,...,g) gear at the j^{th} (j=1,2,...,t) time point (say year) and corresponding effort is expressed as X_{ij}.

To calculate the component of standardized fishing effort for the species corresponding to each gear, the proportion of catch in the total catch by each gear for each year and a weighing factor for each gear is required. Following is the step-wise procedure of effort standardisation:

Step1: Calculate $P_{ijk} = \frac{Y_{ijk}}{Y_{ij.}}$, where $Y_{ij.} = \sum_{k=1}^{s} Y_{ijk}$

Step 2: Obtain the mean and variance of P_{ijk} for each gear and for each species

$$\overline{P_{i.k}} = \frac{1}{t}\sum_{j=1}^{t} P_{ijk} \text{ and } \sigma_{i.k}^2 = \frac{1}{t}\sum_{k=1}^{s}(P_{ijk} - \overline{P_{i.k}})^2$$

Step 3: Calculate weighting factor as

$$W_{i.k} = \frac{\overline{P_{i.k}}}{(\sigma_{i.k}^2 + 1)} \text{ and } W'_{i.k} = \frac{W_{i.k}}{\sum_{i=1}^{g} W_{i.k}}$$

The weighing factor is then adjusted for unit sum. The decomposition of fishing effort for the species is then obtained by multiplying the corresponding total fishing effort for the gear in the year with the proportion of the species for the year corresponding to the same gear and the weighing factor.

Step 4: Obtain the standardized gear-wise fishing effort as

$$E_{ijk} = W'_{i.k} \times P_{ijk} \times X_{ij}$$

Here, the sum of all the gear efforts would give a total effort. But, the efficiency of gears varies so also the capability to catch in an hour which demands scaling the fishing efforts into a single scale. Hence, it is better to express all gears in terms of a single gear (may be the least efficient or the most efficient) by deriving a suitable multiplication factor for each fishing gear.

Step 5: Calculate the catch per unit effort (gear-wise) as

$$CP_{ij} = \frac{Y_{ij.}}{X_{ij}} \text{ and } \overline{CP}_{i.} = \frac{CP_{ij}}{t}$$

The multiplication factor is $\overline{CP}_{i''.} = \frac{\overline{CP}_{i.}}{\overline{CP}_{i'.}}$, where $\overline{CP}_{i'.}$ is the least efficient or the most efficient gear

Step 6: Obtain the standardized fishing effort for k^{th} species at j^{th} time point as

$$\sum_{i=1}^{g} E_{ijk} \times \overline{CP}_{i''.}$$

3. ASSESSMENT OF MARINE STOCK: MONTE CARLO METHOD AND BAYESIAN APPROACH

After obtaining the standardized fishing effort, may be a proxy for CPUE, the stock assessment has been made using the following approaches:

Case 1: *when a measure of fishing effort is available*

Case 2: *when fishing effort is not available (data poor situation)*

Case1 is based on the delay difference model to describe nonlinear population dynamics. State–space model allows incorporation of random errors in both the biomass dynamics equations and the observations. Because the biomass dynamics are nonlinear, the common Kalman filter is generally not applicable for parameter estimation. However, it is demonstrated by (Miller and Meyer, 1998) that the Bayesian approach can handle any form of nonlinear relationship in the state and observation equations as well as realistic distributional assumptions. Difficulties with posterior calculations are overcome by the Gibbs sampler in conjunction with the adaptive rejection Metropolis sampling algorithm (Millar and Meyer, 1998; Froese *et al.* 2017). This approach has been named as BSM and fitted to catch and standardised fishing effort data.

CMSY estimates biomass, exploitation rate, MSY and related fisheries reference points from catch data and resilience of the species. A prior estimate for biomass (B) relative to carrying capacity (k) i.e. B/k has to be given. Next probable ranges for the maximum intrinsic rate of population increase (r) and carrying capacity (k) are given as inputs which then are filtered with a Monte Carlo approach to detect 'viable' r-k pairs. An R package named *R2jags* (Yu-Sung and Masanao, 2015) was used for sampling the probability distributions of the parameters with the Markov chain Monte Carlo method. This package provides wrapper functions to implement Bayesian analysis in JAGS (Plummer, 2003). The convergence of MCMC model is assessed using Rubin and Gelman Rhat statistics, automatically running a MCMC model till it converges, and implementing parallel processing (using *doparallel* package in R) of a MCMC model for multiple chains. The r-ranges for the species under assessment, the proxies for resilience of the species as provided in FishBase (Froese *et al.,* 2000; Froese and Pauly, 2015) and then converted as given by Froese *et al.* (2017).

Even though we have the standardised fishing effort, case 2 has been used to compare the estimate of MSY and the model parameters.

Both the approaches were implemented using R studio (https://www.rstudio.com/). The inputs of time series of catch and information on species resilience are required for running the code and generate the outputs. In order to run the code, the R-libraries required are *R2jags, coda, lattice, parallel, foreach, doParallel,* and *gplots*.

4. DATA DESCRIPTION

Indian mackerel, *Rastrelliger kanagurta*, is an important pelagic fish resource of Andhra Pradesh. The resource is assumed to exist as a single stock along the coastline of Andhra Pradesh (A.P.). The coastline of Andhra Pradesh, which is 974 kilometers long is spread over nine coastal districts viz., Srikakulam, Vizianagaram, Visakhapatnam, East Godavari, West Godavari, Krishna, Guntur, Prakasam and Nellore (FRAD, 2018). Several gears have been found to harvest mackerel almost throughout the year. Like any other tropical pelagic fish, mackerel also exhibited seasonal and annual fluctuations in landings.

Indian mackerel catches in A.P. have been reported from various gears viz., mechanized gillnet (MGN), non-mechanized gears (NM), outboard gillnet (OBGN), outboard ringseine (OBRS), outboard trawlnet (OBTN) and mechanized trawlnet including multiday trawlnet (MTN) and some minor gears.

The mackerel landing was estimated from the commercial landings along the coast of A.P. using a scientifically planned sampling design based on a stratified multi-stage random sampling technique (Sukhatme, 1958 and Srinath *et al.*, 2005), where the stratification is done over space and time. Time series of catch and effort (in hours of operation) from 1997 to 2018 taken from National Marine Fishery Resources Data Centre (NMFDC) of CMFRI, Kochi have been used for the analysis. Standardised fishing effort has been estimated using the proposed method.

5. RESULTS AND DISCUSSION

The annual landings of Indian mackerel in Andhra Pradesh ranged from a low of 6418t (2007) to a high of 55813t (2014) during the study period (Fig. 1a) with an average annual landing of 20551t (SD = 10216). Mackerel landings showed an increasing but variable trend from 1997 onwards, reaching the peak in 2014 and then showed a declining trend. Motorized ring seines (OBRS) landed the highest quantity of Indian mackerel along AP coast during the study period (Table 1). Besides, the summary of fishing effort exerted by major gears in terms of Actual Fishing Hours (in 1000 hrs) has also been given in Table 1. MTN is the gear which operated for a maximum of 3982 (SD = 1212) and OBTN with minimum of 148 (SD = 104).

Table 1. Average landing (in tonnes) and Actual Fishing Hours by each major gear

Gears	Mackerel Landing (t)	AFH ('000 hrs)
MTN	6428, n=22 (2827)	3982 (892)
MGN	1123, n=14 (1412)	438 (338)
NM	5100, n=22 (2711)	2565 (1212)
OBGN	4583, n=22 (2311)	2125 (475)
OBRS	10135, n=10 (6545)	184 (96)
OBTN	208, n=9 (456)	148 (104)

n=number of years with mackerel catch; Standard deviation in parenthesis

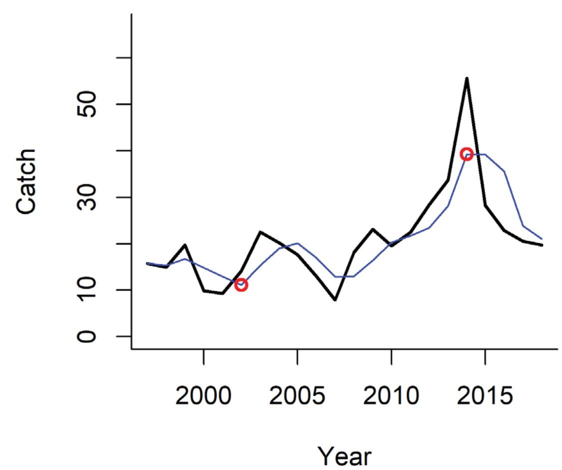


Fig. 1a. Time series of Indian mackerel landings during 1997-2018 (Blue line is the three year moving average, maximum and minimum landings are denoted with red dots)

The standardised fishing effort using the proposed method during the study period indicated an increasing trend with maximum fishing effort exerted in 2015 (Fig. 1b).

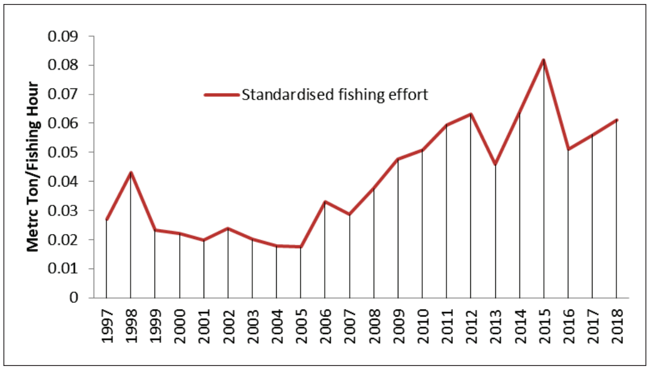


Fig. 1b. Time series of standardised fishing effort during 1997-2018

Since mackerel landings during the initial and final period of the time series are less, lower prior value for B/k was thought to be reasonable. Thus the prior ranges for B/k in the initial and final year were set to 0.2-0.6. Since mackerel landings during the intermediate period were high, the prior range for B/k was set to 0.5 - 0.9 for this period.

FishBase (Froese *et al.*, 2000; Froese and Pauly, 2015) has provided the proxies for resilience of various fish resources and used to set the prior r-ranges by converting as (0.6 – 1.5 for High; 0.2 – 0.8 for Medium; 0.05 – 0.5 for Low and 0.015 – 0.1 for Very low) given by Froese *et al.* (2017). Prior ranges for q are obtained as follows:

$$q_{low} = \frac{0.25 r_{pgm} CPUE_{mean}}{C_{mean}} \text{ and } q_{low} = \frac{0.5 r_{high} CPUE_{mean}}{C_{mean}}$$

where q_{low} is the lower prior for the catchability coefficient for stocks with high recent biomass, r_{pgm} is the geometric mean of the prior range for r, $CPUE_{mean}$ is the mean of catch per unit effort over the last 5 or 10 years, and C_{mean} is the mean catch over the same period. where q_{high} is the upper prior for the catchability coefficient for stocks with high recent biomass, r_{high} is the upper prior range for r. Prior ranges for r, k and q are 0.2-0.9, 43.6-785 and 4.19e-07 - 1.78e-06 respectively.

Once the prior values were given as inputs along with the landings data, the next step in the analysis is to search for viable r-k pairs (Fig.2). Grey colour indicates the viable r-k pairs that fulfilled the CMSY conditions.

The most probable r-k pair is marked by the blue cross, with indication of approximate 95% confidence limits. The black dots show the estimates of the BSM method, with the red cross indicating the 95% confidence limits.

Here, the resilience range of r = 0.2 to 0.9 seems to be meaningful as the points show convergence and fewer viable r-k combinations are found at the end of the r range. It also showed a close agreement with estimated r-k by both the approaches.

Once the r-k pair was selected the relative biomass along with confidence limits was predicted by both the CMSY and BSM method (Fig. 3). The bold curve (blue colour) in Fig.3 is the relative biomass predicted by CMSY, with confidence limits (dotted curves). The normal curve (red colour) indicates CPUE scaled by the catchability coefficient estimated by BSM. The horizontal dashed line indicates biomass at MSY (B_{msy}) and the dotted line indicates half of B_{msy}.

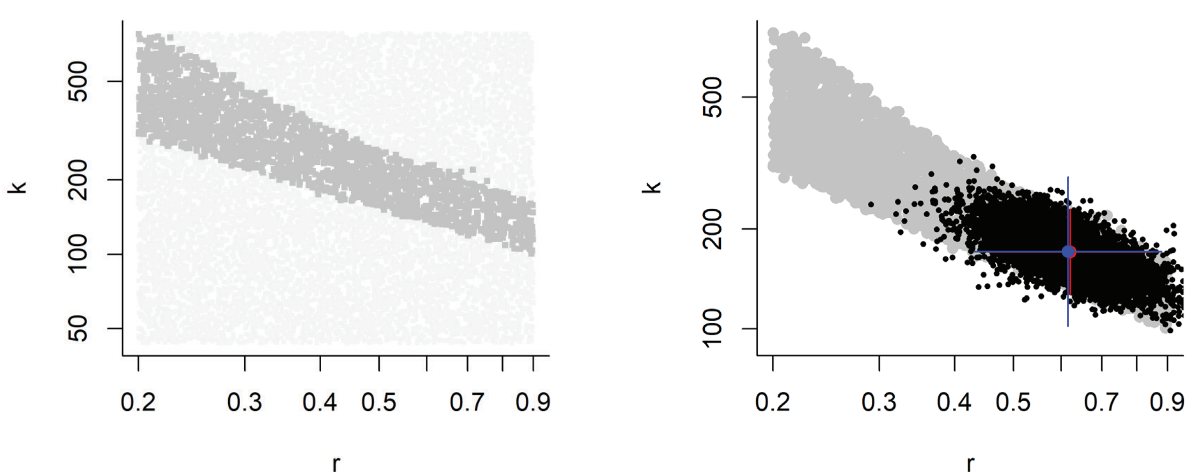


Fig. 2. Search for viable r-k pairs

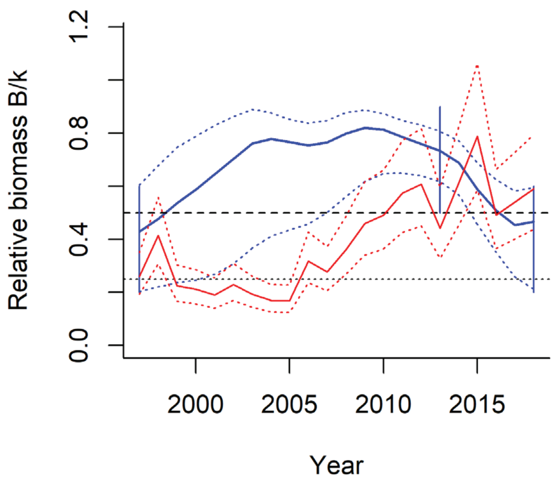


Fig. 3. Relative biomass

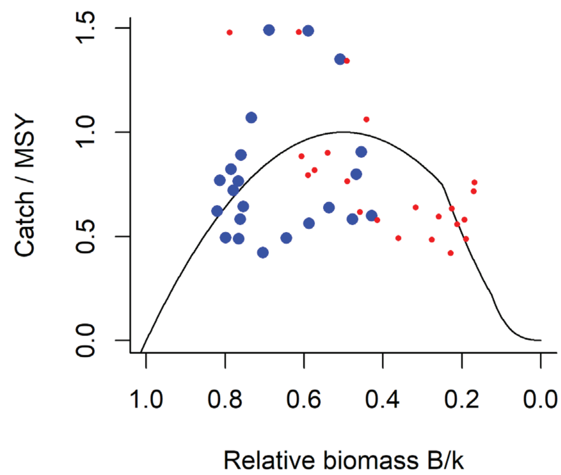


Fig. 4. The ratio of catch to MSY and relative biomass (*B/k*) over years

The relative biomass plot indicated that both in the initial and final years the biomass in relation to carrying capacity was low. This result follows based on the prior estimates of *B*/*k* that we had given. The intervening years showed a high relative biomass. The low relative biomass could be a reflectance of the lower yields from the fishery which was operating at lower fishing effort during the initial years of the study period. From 2005 onwards the fishing effort has been steadily increasing which has also resulted in higher landings since 2005. During this period the relative biomass was above MSY levels. However the relative biomass fell below MSY by 2015 indicating that the stock of Indian mackerel along AP coast is overfished. The overfished status of Indian mackerel along AP coast is further highlighted in the CMSY/BSM output showing catch relative to MSY over biomass relative to unexploited stock size (Fig.4). The red dots indicate estimates by BSM, and the blue dots indicate estimates by CMSY. The indentation of the parabolas below 0.25 *k* (half of B_{msy}) results from the inclusion of a stock–recruitment model which assumes reduced recruitment at low stock sizes.

The points which are above the curve indicate overfishing and shrinking of biomass and the points below the curve indicate sustainable exploitation and growth of the stock. Here, the points are clustered around the equilibrium curve, thus giving confidence in the assessment.

The estimates of MSY and model parameters along with their confidence limits are shown in Table 2. It can be seen from the table that the estimate of MSY is very close by both the approaches with smaller confidence in case of BSM. As BSM takes into account of CPUE, further management plans have been derived based on the BSM results. The landings of Indian mackerel since 2016 has fallen below the estimated MSY.

Table 2. Estimates of MSY and model parameters along with confidence limits

Parameters	CMSY	BSM
MSY	26500 (19200 – 36400)	26600 (20900 - 33800)
r	0.616 (0.431-0.879)	0.623 (0.457-0.848)
k	172000 (102000-289000)	171000 (127000-229000)
Relative biomass in last year (*B*2018/*k*)	0.458 (0.214 - 0.596)	0.532 (0.377 - 0.686)
Exploitation F/(r/2) in last year	0.815	0.7 (0.542 - 0.987)
q	-	6.1e-07 (4.63e-07 - 8.03e-07)
B_{msy}	-	85300 (63500- 115000)
Fishing mortality (F_{msy})	-	0.311 (0.229 - 0.424)
F_{msy} in last year	-	0.218 (0.169 - 0.307)

The plots of landings vs MSY and that of B/B_{msy} (Fig. 5) also indicate the over-fished status of Indian mackerel along AP coast during 2016 onwards. The horizontal dashed line in first plot indicates MSY with lower and upper confidence limit of MSY in grey colour. The bold curve in second plot is the biomass predicted by BSM, with confidence limits (grey colour). The horizontal dashed line indicates B_{msy} and the dotted

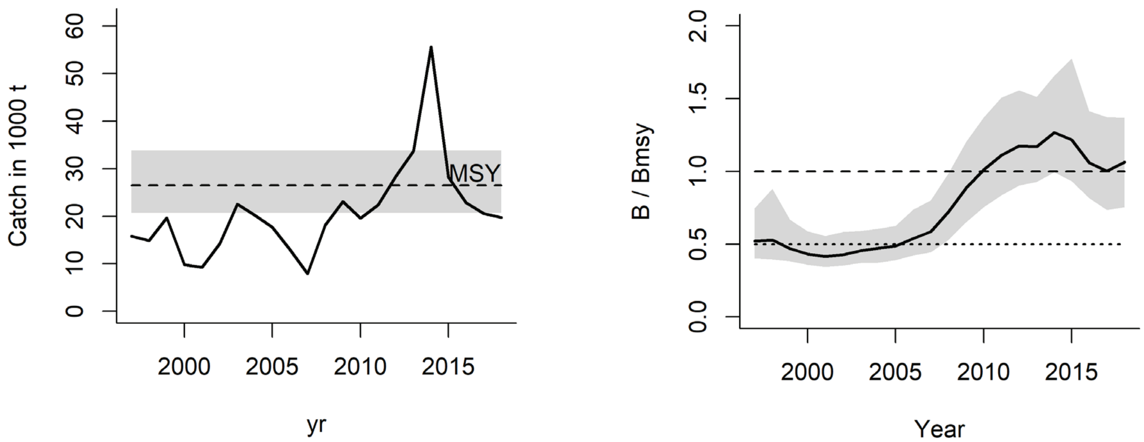


Fig. 5. Catch in comparison to MSY and (B/B_{msy}) over years

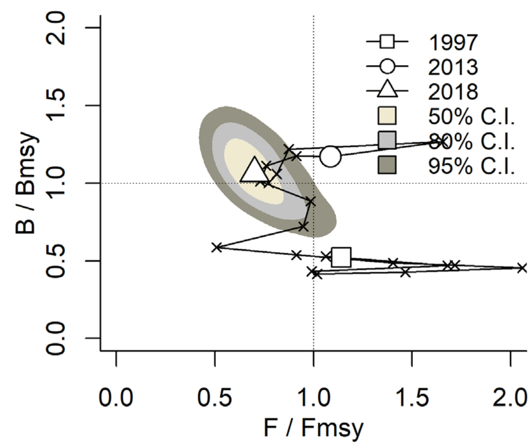


Fig. 6. Development of biomass and exploitation relative to B_{msy} (horizontal dashed line) and F_{msy} (vertical dashed line) for Indian mackerel along AP coast

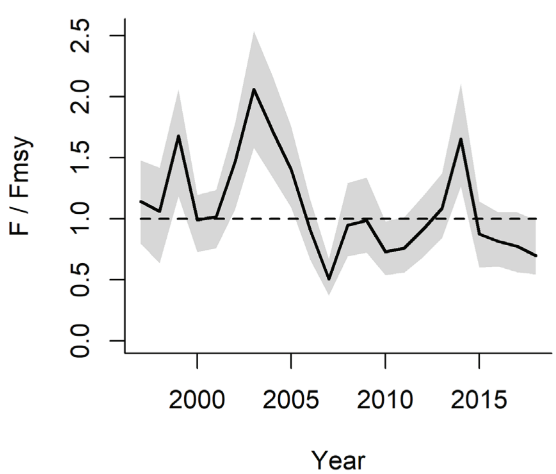


Fig. 7. F/F_{msy} over time for Indian mackerel along AP coast

line indicates half of B_{msy}. The solid line is just above the B_{msy} line during the last two years indicating that current biomass is slightly more than biomass at MSY. Ideally this ratio should be as high as possible. Levels near to 1 indicate that the biomass of the stock of Indian mackerel along AP coast is just at the threshold of being unhealthy.

The plots of current fishing mortality (F) in relation to F at MSY (F_{msy}) (Fig. 6 and 7) indicated that the current fishing mortality is lower than fishing mortality at MSY. However, since current biomass is almost the same level as B_{msy} the stock can be thought to be almost at the edge of unsustainable fishing.

Thus from the above results it can be inferred that the rate of exploitation has been highly fluctuating over period. The current level of exploitation is low as compared to earlier years. Biomass which had been high in intermediate years has declined beyond 2016 due to high exploitation in the intermediate years. The present scenario indicates that a management plan for Indian mackerel along A.P. is needed to ensure its sustainable utilization and that the confidence limit of MSY can serve as guidance for fixing catch limits.

6. CONCLUSIONS

Stock assessment of individual species becomes difficult when a species is targeted by various gears and each gear may harvest more than the species targeted. Since the capacity of the gears vary and also each gear may contain multiple species, the effort made to catch a resource cannot be considered as the sum of duration of operation of all the gears. Here, a new methodology for the standardization of fishing efforts and assessing the stock status of Indian Mackerel (*Rastrelliger kanagurta*) using Bayesian state-space implementation of the Schaefer production model has been discussed. A Monte Carlo method for estimating fisheries reference points from catch using species resilience has also used to assess the stock status in the absence of biomass/ CPUE.

ACKNOWLEDGEMENT

Authors express their sincere thanks to the anonymous referee for the comments and suggestions which helped in improving the quality of the manuscript. Authors thank Professor Rainer Froese, GEOMAR Helmholtz Centre for Ocean Research Kiel, D€usternbrooker Weg 20, 24105 Kiel, Germany for his constructive discussions on the use of CMSY and BSM for stock assessment. We also thank Dr Jan Michael Vakily, Team Leader of the Coastal and Marine Protected Areas Team, Indo-German Biodiversity Programme of GIZ India, Dr. Maria Lourdes D. Palomares (Sea Around Us, University of British Columbia, Canada) and Ms Crispina Binohlan, FIN, Philippines for educating us on CMSY and BSM and also for making the R codes available to us. We thank the Director, CMFRI for providing necessary facilities to carry out the research. Authors thank the organizers of 72nd ISAS for selecting this paper for the best paper award.

REFERENCES

Daniel K., Gibson-Reinemer, Brian S. Ickes, John H. Chick (2016) Development and assessment of a new method for combining catch per unit effort data from different fish sampling gears: Multi-Gear Mean Standardization (MGMS), *Canadian Journal of Fisheries and Aquatic Science,* **74(1)**, 8-14.

FRAD, CMFRI (2018). Marine Fish Landings in India-2017, Technical Report, CMFRI, Kochi.

Froese, R. and Pauly, D. eds (2015) FishBase. World Wide Web electronic publication. www.fishbase.org, version 524 © 2016 John Wiley & Sons Ltd.

Froese, R., Demirel, N., Coro, G., Kleisner, K.M. and Winker, H. (2017). Estimating fisheries reference points from catch and resilience, *Fish and Fisheries*, **18**, 506-526

Froese, R., Palomares, M.L.D. and Pauly, D. (2000) Estimation of life-history key facts. In: FishBase 2000: Concepts, Design and Data Sources (eds R. Froese and D. Pauly), ICLARM, Philippines, p. 167-175; 344 p

Graham, M. (1935). Modern theory of exploiting a fishery, and application to North Sea trawling. *Journal de Conseil International pour l'Exploration de la Mer*, **10**, 264-274.

Hilborn, R.; Walters, C.J. (1992). Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty. Chapman and Hall. 587 p.

Millar, R.B. and Meyer, R. (1998) Nonlinear state-space modeling of fisheries biomass dynamics using metropolis-hastings within gibbs sampling. Technical Report STAT9901. Department of Statistics, University of Auckland, Auckland, New Zealand, 33 p.

Plummer, M. (2003) JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In: *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)* March 20-22, Vienna, 2003. (eds K. Hornik, F. Leisch and A. Zeileis), Vienna Technical University, Vienna pp. 20-22.

Rochman, F., Setyadji, B. and Wujdi, A. (2017). Standardizing CPUE of albacore in eastern Indian Ocean, *Indonesian Fisheries Research Journal*, **23(1)**, 29-38.

Setyadji, B., Andrade, H.A. and Proctor, C.H. (2018). Standardization of catch per unit effort with high proportion of zero catches: an application to black marlin *Istiompax indica* (cuvier, 1832) caught by the indonesian tuna longline fleet in the eastern indian ocean, *Turish. Journal of Fisheries & Aquatic Science*, **19(2)**, 119-129

Sparre, P. and S.C. Venema. (1992). Introduction to tropical fish stock assessment. *FAO Fisheries Technical Paper*, 306/1: 376 pp.

Srinath, M. Kuriakose, S. and Mini, K.G. (2005). Methodology for the estimation of marine fish landings in India, CMFRI Special Publication No. **86**, p.57

Stamatopoulos, C. and Abdallah, M. (2015). Standardization of fishing effort in Qatar fisheries: methodology and case studies. *Journal of Marine Science: Research and Development* 5:170. doi:10.4172/2155-9910.1000170

Sukhatme, P.V. (1958) Sampling technique for estimating the catch of sea fish in India, *Biometrics*, **14(1)**, 78-96

Vivekanandan, E. (2005). Stock assessment of tropical marine fishes. ICAR, New Delhi, 115 p.

Yu-Sung, S. and Masanao, Y. (2015). R2jags: Providing wrapper functions to implement Bayesian analysis in JAGS. https://cran.r-project.org/web/packages/R2jags/index.html

Available online at www.isas.org.in/jisas
Journal of the Indian Society of Agricultural Statistics 74(1) 2020 41–50

New Series of Optimal Covariate Designs in CRD and RBD set-ups

Hiranmoy Das[1], Anurup Majumder[2], Manoj Kumar[3] and Dikeshwar Nishad[4]

[1]ICAR- Indian Institute of Soil Science, Bhopal
[2]Bidhan Chandra Krishi Viswavidyalaya, Nadia
[3]ICAR-Central Institute of Agricultural Engineering, Bhopal
[4]Pt. Shiv Kumar Shastri College of Agriculture and Research Station, Rajnandgaon

Received 22 April 2019; Revised 30 August 2019; Accepted 04 January 2020

SUMMARY

The study of optimal design for covariate models in CRD set-up was initiated by Troya (1982a, 1982b). Das *et al.* (2003) followed up the study and extended for RBD set-up. Recently Das *et al.* (2015) published a book on 'Optimal Covariate Designs'. In the present study, one new series of global optimal covariate designs in CRD set-up and two new series in RBD set-up have been developed. The new OCDs in CRD or RBD designs require only two Hadamard matrices of order 2 and 4. The developed global optimal covariate designs in CRD set-up have v (=0; mod 4 or =2; mod 4) number of treatments, and the developed first series of global optimal covariate designs in RBD set-up have the treatment number **v** (=0; mod 4) for any even number of replications or blocks, **b** and do not dependent on the existence of $\mathbf{H}_v$ and $\mathbf{H}_b$. The second series of global optimal covariate designs in RBD set-up require only the existence of $\mathbf{H}_v$. The paper is enriched with examples of optimal covariates. All the developed optimal covariate designs in the present article are not available in the existing literature.

Keywords: ANCOVA, Hadamard Matrix, Optimal Covariate Design (OCD), CRD, RBD.

1. INTRODUCTION

Analysis of Covariance or 'ANCOVA' is a known method by which the error affecting the treatment comparisons may be minimized. The experimental results can be improved by suitably classifying or reclassifying the existing experimental units through a study of the associated covariates or by first suitably choosing the covariate values from a larger lot and then identifying the associated experimental units from a larger pool. The choice or selection of experimental units with suitably defined values of the covariates for a particular design set-up so as to attain the minimum variance or maximum precision for estimating the regression parameters has fascinated the interest of statisticians for the last three decades or a little more. Several authors, like Harville (1974, 1975), Haggstrom (1975) and Wu (1981) had studied the ANCOVA models on the problems of inference on varietal contrasts corresponding to qualitative factors. But the problem of determining the optimum designs for the estimation of regression parameters corresponding to controllable covariates was not a topic of research for many years. Troya (1982a and 1982b) was the pioneer in history in the topic of optimal covariate designs (OCDs) but she restricted to only Completely Randomized Design (CRD) set-up. After a long gap, Das *et al.* (2003) extended the work on OCDs to the block design set-up, viz., Randomized Block Design (RBD) and some series of Balanced Incomplete Block Design (BIBD). They also constructed OCDs for the estimation of covariate parameters. Rao *et al.* (2003) also revisited the problem in CRD and RBD set-ups. They identified that the solutions of construction of OCDs by using Mixed Orthogonal Arrays (MOAs) and thereby giving further insights and some new solutions. Dutta (2004) developed OCDs for BIBDs obtained through Bose's Difference Technique. He utilized conveniently different combinatorial arrangements and

Corresponding author: Hiranmoy Das
E-mail address: hiranmoydas.stat@gmail.com

tools such as Hadamard matrices and different kinds of products of matrices viz., Kronecker product to construct OCDs with as many covariates as possible. Dey and Mukherjee (2006) studied the problem of finding D-optimal designs in the presence of a number of covariates in the one-way set-up. They actually given an upper bound to the determinant of information matrix obtained through diagonal **C**-matrices. Dutta *et al.* (2007) studied the optimum choice of covariates for a series of balanced incomplete block designs (BIBDs). In case of incomplete block designs, the choice of the values of the covariates depends heavily on the allocation of treatments to the plots of blocks; more specifically on the method of construction of the incomplete block design. Based on this, they considered the situation where the block design is a member of the complementary series of balanced incomplete block design (BIBD) with parameters $b = v = s^N+s^{N-1}+\ldots+s+1$, $r = k = s^N$, $\lambda=s^N-s^{N-1}$ of symmetric balanced incomplete block design (SBIBD) obtained through projective geometry. Sinha (2009) gave the solution to accommodate maximum number of covariates in an optimal manner through combinatorially for the standard design layouts such as CRD, RBD, LSD and BIBD. Dutta *et al.* (2010b) considered the problem that when $n\neq 0 \pmod 4$, it is impossible to find designs attaining minimum variance for estimated covariate parameters. In this situation, they considered instead of using the criterion of attaining the lower bound (viz., σ^2/n) to the variance of each of the estimated covariate parameters γ, they found optimum designs with respect to covariate effects using D-optimality criterion retaining orthogonality with respect to treatment and block effect contrasts, where $n=2 \pmod 4$). Dutta *et al.* (2014) extended the work of Dey and Mukherjee (2006) in the sense that for fixed replication numbers of each treatment, an alternative upper bound to the determinant of information matrix has been found through completely symmetric **C**-matrices for the regression coefficients and this upper bound includes the upper bound given in Dey and Mukherjee (2006). Recently, Das *et al.* (2015) has published a book, viz., 'Optimal Covariate Designs' with detail discussion on the topic. Mostly the designs developed by above mentioned authors are global optimal but the development of designs are dependent on existence of Hadamard matrix of order either v or b or k (v be the treatment numbers, b be the number of replications/ blocks in CRD/ RBD and k be the size of blocks in a variance balanced incomplete block design).

In the present piece of investigation, an effort has been made to construct global optimal covariate designs in CRD and RBD set-ups when Hadamard matrices of order $\mathbf{H}_v$ and $\mathbf{H}_b$ do not exist. The study contains five sections including the present introductory section. In section 2, the definition and properties of Special Array are presented. Section 3 and 4 describe the basic models, situations and conditions of the optimal covariate designs (OCDs) for CRD and RBD set-ups, respectively. Construction of a new series of global optimal covariate designs in CRD set-up has also been presented in section 3. Similarly, construction of two new series of global optimal covariate designs in RBD set-up has been given in section 4. Conclusion of the study has been given in section 5.

2. SPECIAL ARRAY; DEFINITION, PROPERTIES

2.1 Definition

A square matrix with elements 1, -1 and 0 of order h having r (≥ 1) number of rows (and columns) with all elements 0 and all the distinct row or column vectors except r rows (or columns) of the matrix are mutually orthogonal will be referred to as **Special Array** (SA) of order h. In SA, each row or column sum is zero except the first row or column. The simplest examples, one for order 3 and two for order 5 are given below:

$$\underset{r=1}{\begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & -1 \end{pmatrix}}, \underset{r=3}{\begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 \end{pmatrix}} \text{ and } \underset{r=1}{\begin{pmatrix} 1 & 1 & 0 & 1 & 1 \\ 1 & -1 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & -1 & -1 \\ 1 & -1 & 0 & 1 & -1 \end{pmatrix}}$$

2.2 Properties

Let the Special Array (SA) of order h be denoted as H_h^*, then

1) $\left|\det H_h^*\right| = 0$; when $r \geq 1$; with $r = 0$, H_h^* becomes a Hadamard Matrix.
2) $H_h^* H_h^{*T} = H_h^{*T} H_h^*$
3) Let H_1^* and H_2^* be two SA of order h_1 and h_2, respectively. Then the Kronecker product of H_1^* and H_2^* is also a SA of order h_1h_2.

3. OCDS IN CRD SET-UP

Let there be v treatments and c covariates in a design with total n experimental units. In matrix notation the model can be represented as

$$(\mathbf{Y}, \mathbf{X\tau} + \mathbf{Z\gamma}, \sigma^2\mathbf{I}_n) \tag{3.1}$$

where, for $1 \leq i \leq v$, $1 \leq j \leq n_i$ (n_i is the number of times the i^{th} treatment is replicated; clearly $\sum_{i=1}^{v} n_i = n$) and $1 \leq t \leq c$, **Y** is an observation vector and **X** is the design matrix corresponding to vector of treatment effects $\boldsymbol{\tau}^{vx1}$ and $\mathbf{Z}^{nxc} = ((z_{ij}^{(t)}))$ is the design matrix corresponding to vector of covariate effects $\boldsymbol{\gamma}^{cx1} = (\gamma_1, \gamma_2, \ldots, \gamma_c)'$. This is referred to as one–way model with covariates without general mean. In the above, **Z** is called covariate matrix of c covariates $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_c$. Here **z**'s are assumed to be controllable non-stochastic covariates. The n values $z_{i1}, z_{i2}, \ldots, z_{in}$ are assumed by the ith covariate $\mathbf{z}_i$ are such that they belong to a finite interval $[a_i, b_i]$ for each i and j, i.e.

$$a_i \leq z_{ij} \leq b_i \tag{3.2}$$

$$\text{i.e. } z_{ij} = \frac{a_i + b_i}{2} + \frac{b_i - a_i}{2} z_{ij}^* \tag{3.3}$$

so that z_{ij}^* lies in [-1, 1] for each i, j. Then replacing z_{ij} by z_{ij}^*'s, we get the same covariate model in a reparametrized scenario. So, without loss of generality, the covariate values z_{ij}'s to vary within [-1, 1].The information matrix with respect to model (3.1) is given by,

$$\sigma^{-2}\mathrm{I}(\eta) = \begin{pmatrix} X'X & X'Z \\ Z'X & Z'Z \end{pmatrix} \text{ where, } \eta' = (\tau', \gamma') \tag{3.4}$$

The problem is to suggest an optimal allocation scheme (for given design parameters n, v, c) for efficient estimation of the treatment effects and the covariate effects by ascertaining the values of the covariates for each one of them, assuming that each one is controllable and quantitative within a stipulated finite closed interval. The information matrix of γ is given by,

$$\sigma^{-2}\mathrm{I}(\gamma) = Z'Z - Z'X(X'X)^{-}X'Z \tag{3.5}$$

where, $(\mathbf{X'X})^{-}$ is a generalized inverse of $\mathbf{X'X}$. According to Rao (1973), $\mathbf{Z'X}(\mathbf{X'X})^{-}\mathbf{X'Z}$ is a positive semi-definite matrix. So from (3.5), it follows that

$$\sigma^{-2}\mathrm{I}(\gamma) \leq Z'Z \tag{3.6}$$

Equality in (3.6) is attained whenever $\mathbf{X'Z} = \mathbf{0}$ (3.7)

If **Z** satisfies (3.7), then treatment effects and covariate effects are orthogonally estimated. In addition, the information matrix $\mathbf{I}(\gamma)$ reduces to $\mathbf{I}(\gamma) = \mathbf{Z'Z}$. The z-values are so chosen that $\mathbf{Z'Z}$ is positive definite, so that from (3.6)

$$Var(\hat{\gamma}_t) \geq \frac{\sigma^2}{\sum_{i=1}^{v}\sum_{j=1}^{n_i} z_{ij}^{(t)2}} \geq \frac{\sigma^2}{n} \tag{3.8}$$

as

$$z_{ij}^{(t)} \in [-1, 1]; \forall i, j, t$$

Now equality in (3.8) holds for all i if and only if the **Z**-matrix is such that

$$\mathbf{z}^{(s)\prime}\mathbf{z}^{(t)} = 0 \text{ for all } s \neq t \tag{3.9}$$

$$\text{and } z_{ij}^{(t)} = \pm 1 \tag{3.10}$$

Condition (3.7) implies that the estimators of ANOVA effects parameters or parametric contrasts do not interfere with those of the covariate effects and conditions (3.9) and (3.10) imply that the estimators of each of the covariate effects are such that these are pair wise uncorrelated, attaining the minimum possible variance. Thus, the covariate effects are estimated with the maximum efficiency if and only if

$$\mathbf{Z'Z} = n\mathbf{I}_c \tag{3.11}$$

along with (3.7). The designs allowing the estimators with the minimum variance are called globally optimal designs (Shah and Sinha, 1989) or optimal covariate design, to be abbreviated as OCD.

Visualizing the **Z**-matrix in a particular design set up satisfying conditions (3.7) and (3.11) is somewhat difficult. In the set-up of the model (3.1), it transpires from Troya Lopes (1982a) that optimal estimation of the treatment effects and the covariates effects is possible when the treatment replications are all necessarily equal, assuming that n is a multiple of v, the number of treatments. Set n = bv, where b is the common replication of treatments. Das *et al.* (2003) had represented each column of the **Z**-matrix by a v×b matrix, viz., **W** with elements of ±1. Condition (3.7) implies that the sum of each row of **W** should be zero. Further, condition (3.11) implies that the sum of products of the corresponding elements i.e. the Hadamard product of $\mathbf{W}^{(s)}$ and $\mathbf{W}^{(t)}$, should also be zero, $1 \leq s < t \leq c$. For orthogonality of sth and tth columns of **Z**, it is required that

$$\sum_{i=1}^{v}\sum_{j=1}^{b} w_{ij}^{(s)} w_{ij}^{(t)} = 0 \tag{3.12}$$

In this case the ANOVA parameters as well as the covariate effect-parameters can be estimated orthogonally and/or most efficiently. This holds simultaneously for c covariates and one can deduce maximum possible value of c for this to happen. As already mentioned, the most efficient estimation of γ-components is possible when conditions (3.7) and (3.11) are simultaneously satisfied and these conditions reduce, in terms of **W**-matrices defined in above, to $\mathbf{C}_1$ and $\mathbf{C}_2$, where

$\mathbf{C}_1$: Each of the c **W**-matrices has all row-sums equal to zero;

$\mathbf{C}_2$**:** The grand total of all the entries in the Hadamard product of any two distinct **W**-matrices reduces to zero.

3.1 Construction of optimum W-matrices for covariate model in CRD set-up:

Definition 3.1: With respect to model (3.1), the c number of **W**-matrices corresponding to the c covariates are said to be optimum if they satisfy the conditions $\mathbf{C}_1$ and $\mathbf{C}_2$ simultaneously.

Under the realization of $\mathbf{C}_1$ and $\mathbf{C}_2$ in terms of optimum **W** matrices, we can develop the following theorem.

Theorem 3.1: If both v and b be two even numbers, then there exists c (=2) optimum covariates in a CRD set-up with v (=0; mod 4 or =2; mod 4) treatments with b replications for each treatment even if $\mathbf{H}_v$ and $\mathbf{H}_b$ do not exist.

Proof (by construction)**:** For construction of optimum **W** matrices of order vxb, we follow the steps given below.

Step 1. Let us consider two Hadamard matrices $\mathbf{H}_4$ and $\mathbf{H}_2$.

$$H_4 = (1, h_1, h_2, h_3) = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \end{pmatrix}, \quad H_2 = (1, h_1^*) = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

Step 2. Using $\mathbf{H}_4$ and $\mathbf{H}_2$, we construct the following three **W*** matrices ($\mathbf{W}_1$*, $\mathbf{W}_2$* and $\mathbf{W}_3$*) of order 4x2 by Kronecker product of the columns (with zero sums) of these two matrices.

$$W_1^* = h_1 \otimes h_1^{*'} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \\ 1 & -1 \\ -1 & 1 \end{pmatrix}, \quad W_2^* = h_2 \otimes h_1^{*'} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \\ -1 & 1 \\ 1 & -1 \end{pmatrix},$$

$$W_3^* = h_3 \otimes h_1^{*'} = \begin{pmatrix} 1 & -1 \\ 1 & -1 \\ -1 & 1 \\ -1 & 1 \end{pmatrix}$$

Step 3. Firstly, repeat each of the $\mathbf{W}_i$* (i = 1, 2, 3) vertically side by side q-1 (≥1) times such that b = 2q. Let the newly matrix be denoted as $\mathbf{W}_i$** (i = 1, 2, 3) of order 4xb.

$$W_1^{**} = \left(\begin{pmatrix} 1 & -1 \\ -1 & 1 \\ 1 & -1 \\ -1 & 1 \end{pmatrix} \overbrace{\begin{pmatrix} 1 & -1 \\ -1 & 1 \\ 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -1 & 1 \\ 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{matrix} \ldots \\ \ldots \\ \ldots \\ \ldots \end{matrix} \begin{pmatrix} 1 & -1 \\ -1 & 1 \\ 1 & -1 \\ -1 & 1 \end{pmatrix}}^{q-1} \right)$$

Similarly, construct the W_2** and W_3**.

Step 4. Next, repeat the first pair of rows of each of the $\mathbf{W}_i$** matrix horizontally p-2 (≥1) times such that v = 2p. Let the constructed matrix be denoted as $\mathbf{W}_i$ (i = 1, 2, 3) of order vxb.

$$W_1 = \begin{pmatrix} \begin{pmatrix} 1 & -1 & 1 & -1 & 1 & -1 & \ldots & 1 & -1 \\ -1 & 1 & -1 & 1 & -1 & 1 & \ldots & -1 & 1 \end{pmatrix} \\ \begin{matrix} 1 & -1 & 1 & -1 & 1 & -1 & \ldots & 1 & -1 \\ -1 & 1 & -1 & 1 & -1 & 1 & \ldots & -1 & 1 \end{matrix} \\ \left. \begin{matrix} \begin{pmatrix} 1 & -1 & 1 & -1 & 1 & -1 & \ldots & 1 & -1 \\ -1 & 1 & -1 & 1 & -1 & 1 & \ldots & -1 & 1 \end{pmatrix} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ \begin{pmatrix} 1 & -1 & 1 & -1 & 1 & -1 & \ldots & 1 & -1 \\ -1 & 1 & -1 & 1 & -1 & 1 & \ldots & -1 & 1 \end{pmatrix} \end{matrix} \right\} p-2 \end{pmatrix}$$

Similarly, construct the W_2 and W_3 matrices.

Step 5. Among the three **W** matrices, either the pair ($\mathbf{W}_1$, $\mathbf{W}_3$) or **(**$\mathbf{W}_2$, $\mathbf{W}_3$) are satisfying the conditions $\mathbf{C}_1$ and $\mathbf{C}_2$ simultaneously for global optimality.

For easy understanding of the above steps, the following example will be useful.

Example 3.1: Let us consider a CRD with v = 6 and b = 6. The constructional procedure of two optimum **W** matrices is given below:

$$H_4 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \end{pmatrix}, \; H_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

By Kronecker product of the columns (with zero sums) of these two matrices, we get

$$W_1^* = \begin{pmatrix} 1 & -1 \\ -1 & 1 \\ 1 & -1 \\ -1 & 1 \end{pmatrix}, \; W_2^* = \begin{pmatrix} 1 & -1 \\ -1 & 1 \\ -1 & 1 \\ 1 & -1 \end{pmatrix}, \; W_3^* = \begin{pmatrix} 1 & -1 \\ 1 & -1 \\ -1 & 1 \\ -1 & 1 \end{pmatrix}$$

In each $\mathbf{W}_i^*$, (i=1,2,3), the pair of columns replicated vertically twice and we get

$$W_1^{**} = \begin{pmatrix} 1 & -1 & 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 & -1 & 1 \end{pmatrix}, \; W_2^{**} = \begin{pmatrix} 1 & -1 & 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 & -1 & 1 \\ -1 & 1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 \end{pmatrix}, \; W_3^{**} = \begin{pmatrix} 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 & -1 & 1 \\ -1 & 1 & -1 & 1 & -1 & 1 \end{pmatrix}$$

Again in each $\mathbf{W}_i^{**}$, (i=1,2,3), first pair of rows further replicated one time and we get

$$W_1 = \begin{pmatrix} 1 & -1 & 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 & -1 & 1 \end{pmatrix}, \; W_2 = \begin{pmatrix} 1 & -1 & 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 & -1 & 1 \\ -1 & 1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 & -1 & 1 \end{pmatrix}, \; W_3 = \begin{pmatrix} 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 & -1 & 1 \\ -1 & 1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 \end{pmatrix}$$

$\{\mathbf{W}_1, \mathbf{W}_3\}$ and $\{\mathbf{W}_2, \mathbf{W}_3\}$ are the two sets, each having two optimum **W** matrices satisfying the conditions $\mathbf{C}_1$ and $\mathbf{C}_2$ simultaneously.

4. OCDS IN RBD SET-UP

For two-way layout, the set-up can be written as

$$(\mathbf{Y}, \mu\mathbf{1}+\mathbf{X}_1\boldsymbol{\tau}+\mathbf{X}_2\boldsymbol{\beta}+\mathbf{Z}\boldsymbol{\gamma}, \sigma^2\mathbf{I}) \tag{4.1}$$

where μ, as usual, stands for the general effect, $\boldsymbol{\tau}^{vx1}$, $\boldsymbol{\beta}^{bx1}$ represent vectors of treatment and block effects, respectively, $\mathbf{X}_1^{nxv}$ and $\mathbf{X}_2^{nxb}$ are the corresponding incidence matrices, respectively. **Y** and **Z** as usual, represents an observation vector of order nx1 and the design matrix of order nxc corresponding to vector of covariate effects $\boldsymbol{\gamma}^{cx1}$, respectively.

The information matrix for the whole set of parameters $\eta = (\mu, \tau', \beta', \gamma')'$ underlying a design d with $\mathbf{X}_{1d}$, $\mathbf{X}_{2d}$ and $\mathbf{Z}_d$ as the versions of $\mathbf{X}_1$, $\mathbf{X}_2$ and **Z** in (4.1):

$$I_d(\eta) = \begin{pmatrix} n & 1'X_{1d} & 1'X_{2d} & 1'Z_d \\ & X_{1d}'X_{1d} & X_{1d}'X_{2d} & X_{1d}'Z_d \\ & & X_{2d}'X_{2d} & X_{2d}'Z_d \\ & & & Z_d'Z_d \end{pmatrix} \tag{4.2}$$

For the covariates, without loss of generality, the (location scale)-transformed version, $\left|z_{ij}^{(t)}\right| \le 1;\ i, j, t$. From (4.2), it is evident that orthogonal estimation of treatment and block effect contrasts on one hand and covariate effects on the other is possible when the conditions

$$\mathbf{X}_{1d}'\mathbf{Z}_d = \mathbf{0}, \text{ and } \mathbf{X}_{2d}'\mathbf{Z}_d = \mathbf{0} \tag{4.3}$$

are satisfied. It is to be noted that under (4.3), $\mathbf{1}'\mathbf{Z}_d = \mathbf{0}'$ also holds. Further, the most efficient estimation of γ-components is possible whenever, in addition to (4.3), we can also ascertain

$$\mathbf{Z}_d'\mathbf{Z}_d = n\mathbf{I}_c \tag{4.4}$$

For an RBD set-up, following Das *et al.* (2003), we recast each column of the $\mathbf{Z}^{nxc}$ = (±1) matrix by a **W**-matrix of order vxb. Corresponding to the treatment x block classifications, conditions (4.3) and (4.4) reduce, in terms of **W**-matrices, to $\mathbf{C}_1^*$ - $\mathbf{C}_3^*$ where

$\mathbf{C}_1^*$: Each **W**-matrix has all column-sums equal to zero;

$\mathbf{C}_2^*$: Each **W**-matrix has all row-sums equal to zero;

$\mathbf{C}_3^*$: The grand total of all the entries in the Hadamard product of any two distinct **W**-matrices reduces to zero.

4.1 Construction of optimum W-matrices for covariate model in RBD set-up:

Definition 4.1: With respect to model (4.1), the c number of **W**-matrices corresponding to the c covariates are said to be optimum if they satisfy conditions $\mathbf{C}_1^*$, $\mathbf{C}_2^*$ and $\mathbf{C}_3^*$ simultaneously.

Now, under the realization of $\mathbf{C}_1^*$, $\mathbf{C}_2^*$ and $\mathbf{C}_3^*$ in terms of optimum **W** matrices, we can develop the following theorem.

Theorem 4.1: If both v and b be two even numbers, then there exists c (= 3) optimum covariates in a Randomized Complete Block Design (RCBD or RBD) with v (=0; mod 4) treatments and b number of blocks even if $\mathbf{H}_v$ and $\mathbf{H}_b$ do not exist.

Proof (by construction)**:** For construction of optimum W matrices of order vxb, we follow the steps given below.

Step 1. Let us consider two Hadamard matrices $\mathbf{H}_2$ and $\mathbf{H}_4$ as shown in step 1 of theorem 3.1.

Step 2. Using $\mathbf{H}_2$ and $\mathbf{H}_4$, we construct three **W*** matrices ($\mathbf{W}_1$*, $\mathbf{W}_2$* and $\mathbf{W}_3$*) of order 2x4 by Kronecker product of the columns (with zero sums) of these two matrices.

$$W_1^* = h_1 \otimes h_1^{*'} = \begin{pmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \end{pmatrix},\quad W_2^* = h_1 \otimes h_2^{*'} = \begin{pmatrix} 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \end{pmatrix},$$

$$W_3^* = h_1 \otimes h_3^{*'} = \begin{pmatrix} 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \end{pmatrix}$$

Step 3. Firstly, repeat each of the $\mathbf{W}_i$* (i = 1, 2, 3) vertically side by side q-1 (≥1) times such that v = 4q. Let the newly matrix be denoted as $\mathbf{W}_i$** (i = 1, 2, 3) of order 2xv.

$$W_1^{**} = \left(\boxed{\begin{matrix} 1\,{-1}\ 1\,{-1} \\ -1\ 1\,{-1}\ 1 \end{matrix}}\ \overbrace{\boxed{\begin{matrix} 1\,{-1}\ 1\,{-1} \\ -1\ 1\,{-1}\ 1 \end{matrix}}\ \begin{matrix} \cdots \\ \cdots \end{matrix}\ \boxed{\begin{matrix} 1\,{-1}\ 1\,{-1} \\ -1\ 1\,{-1}\ 1 \end{matrix}}}^{q-1} \right)$$

Similarly, construct the W_2** and W_3**.

Step 4. Repeat each of the $\mathbf{W}_i$** matrix horizontally p-1 (≥1) times such that b = 2p. Let the constructed matrix be denoted as $\mathbf{W}_i$ ***(i = 1, 2, 3) of order bxv.

$$W_1^{***} = \left(\begin{matrix} \boxed{\begin{matrix} 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1} \ldots 1\,{-1}\ 1\,{-1} \\ -1\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1 \ldots {-1}\ 1\,{-1}\ 1 \end{matrix}} \\ \left. \begin{matrix} \boxed{\begin{matrix} 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1} \ldots 1\,{-1}\ 1\,{-1} \\ -1\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1 \ldots {-1}\ 1\,{-1}\ 1 \end{matrix}} \\ \cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots \\ \boxed{\begin{matrix} 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1} \ldots 1\,{-1}\ 1\,{-1} \\ -1\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1 \ldots {-1}\ 1\,{-1}\ 1 \end{matrix}} \end{matrix} \right\} p-1 \end{matrix} \right)$$

Similarly, construct the W_2*** and W_3***.

Step 5. After taking the transpose of each $\mathbf{W}_i$*** matrices, we get the set of desired covariate matrices ($\mathbf{W}_1$, $\mathbf{W}_2$ and $\mathbf{W}_3$) satisfying the conditions $\mathbf{C}_1^*$, $\mathbf{C}_2^*$ and $\mathbf{C}_3$* simultaneously for global optimality.

For easy understanding of the above steps, the following example will be useful.

Example 4.1: Let us consider a RBD with v = 20 and b = 6. The method of construction of three **W** matrices are given below:

By taking the Kronecker product of the columns (with zero sums) of $\mathbf{H}_2$ and $\mathbf{H}_4$ matrices, we get,

$$W_1^* = \begin{pmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \end{pmatrix},\quad W_2^* = \begin{pmatrix} 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \end{pmatrix},\quad W_3^* = \begin{pmatrix} 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \end{pmatrix}$$

In each $\mathbf{W}_i$*, (i=1,2,3), whole set of columns replicated vertically four times and we get

$$W_1^{**} = \begin{pmatrix} 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1} \\ -1\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1 \end{pmatrix},$$

$$W_2^{**} = \begin{pmatrix} 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1 \\ -1\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1} \end{pmatrix}$$

$$W_3^{**} = \begin{pmatrix} 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1} \\ -1\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1 \end{pmatrix}$$

Again in each $\mathbf{W}_i$**, (i=1,2,3), the whole set of rows further replicated horizontally two times, then we get,

$$W_1^{***} = \begin{pmatrix} 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1} \\ -1\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1 \\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1} \\ -1\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1 \\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1} \\ -1\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1\,{-1}\ 1 \end{pmatrix},$$

$$W_2^{***} = \begin{pmatrix} 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1 \\ -1\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1} \\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1 \\ -1\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1} \\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1 \\ -1\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1} \end{pmatrix}$$

$$W_3^{***} = \begin{pmatrix} 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1} \\ -1\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1 \\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1} \\ -1\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1 \\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1} \\ -1\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1\,{-1}\,{-1}\ 1\ 1 \end{pmatrix}$$

After transpose of each $\mathbf{W}_i$***, (i=1,2,3), we get the ultimate three optimum **W** matrices satisfying the conditions $\mathbf{C}_1^*$, $\mathbf{C}_2^*$ and $\mathbf{C}_3$* simultaneously, e.g., $W_1 = W_1^{***'}$, $W_2 = W_2^{***'}$ and $W_3 = W_3^{***'}$.

Corollary 4.1: The optimal covariate design in RBD developed by theorem 4.1 is true for CRD with similar v and b.

Proof: Straight forward from the definition of CRD.

Theorem 4.2: The existence of a Hadamard matrix of order v, $\mathbf{H}_v$ and a Special Array of order b, $\mathbf{H}_b$* (b=0; mod 4) with r rows and columns with all zero elements in middle, implies the existence of either (i) $(r-1)^2$ or (r-1)(v-1) optimal covariates when $(r-1)^2$ or (r-1)(v-1) is less than (v-1)(b-r-1) or (ii) (v-1)(b-r-1) optimal covariates when $(r-1)^2$ or (r-1)(v-1)≥(v-1)(b-r-1) of a RBD with v treatments in b blocks provided $\mathbf{H}_r$ and $\mathbf{H}_{b-r}$ exist and r = v/m, where, m is any real valued positive integer number.

Proof (by construction): For construction of optimum W matrices of order vxb, we follow the steps given below.

Step 1. Let us consider a Hadamard matrix of order v, $\mathbf{H}_v=(\mathbf{1}, \mathbf{h}_1, \mathbf{h}_2,\ldots,\mathbf{h}_{v-1})$.

Step 2. Let us construct a Special Array $\mathbf{H}_b$* of order b from $\mathbf{H}_{b-r}$ with r rows and columns with all zero elements in middle, i.e., $(\mathbf{1}^*, \mathbf{h}_1^*, \mathbf{h}_2^*,\ldots, \mathbf{h}_{(b-r)/2-1}^*, \mathbf{0},\ldots, \mathbf{0}, \mathbf{h}_{(b-r)/2}^*,\ldots, \mathbf{h}_{b-r-1}^*)$.

$$H_b^* = \begin{array}{c} \overset{\longleftrightarrow}{r} \\ \begin{pmatrix} 1 & 1 & 0 & \ldots\ldots & 0 & 1 & 1 \\ 1 & -1 & 0 & \ldots\ldots & 0 & -1 & 1 \\ 0 & 0 & 0 & \ldots\ldots & 0 & 0 & 0 \\ \ldots & & & & & & \ldots \\ \ldots & & & & & & \ldots \\ 0 & 0 & 0 & \ldots\ldots & 0 & 0 & 0 \\ 1 & 1 & 0 & \ldots\ldots & 0 & -1 & -1 \\ 1 & -1 & 0 & \ldots\ldots & 0 & 1 & -1 \end{pmatrix} \end{array} \updownarrow r$$

Step 3. Using $\mathbf{H}_b$*and $\mathbf{H}_v$, we get (b-r-1) sets of (v-1) W_{ij}* matrices of order bxv (without considering the first column and r columns with all zeros) by taking the Kronecker product of the columns (with zero sums) of the above matrices, where $i=1,2,\ldots,(b\text{-}r\text{-}1)$ and $j=1,2,\ldots,(v\text{-}1)$. In each of the W_{ij}* matrix there are r rows with all elements zero in the middle.

$W_{ij}^* = h_i^* \otimes h'_j$, $\otimes$ denotes the Kronecker product

Step 4. As $\mathbf{H}_v$ and $\mathbf{H}_r$ both exist, following the Theorem 3.4.1 (Das *et al.*, 2015), we construct orthogonal W** matrices either $(r-1)^2$ numbers of order r or (r-1)(v-1) numbers of order rxv.

Step 5. In each **W*** matrix, insert the first W** matrix of order r in the first r columns of W_1* matrix and replicate the selected W** matrix (v-r)/r times or insert the first W** matrix of order rxv in the r rows with all elements zero in the middle of W_1* matrix, such that all the r rows with all elements zero has been replaced by ±1. Let the resulting matrix be W_1'. Repeat the procedure with other W** matrices in the remaining **W*** matrices till all W** matrices or all W* matrices have been covered totally. So, we get either (i) $(r-1)^2$or (r-1)(v-1) W′ matrices of order bxv when $(r-1)^2$ or (r-1)(v-1) < (b-r-1)(v-1) or (ii) (b-r-1)(v-1) W′ matrices of order bxv when $(r-1)^2$ or (r-1)(v-1) ≥ (b-r-1)(v-1), which are orthogonal to each other and all the **W′** matrix has all column-sums and row-sums equal to zero. Finally, the desired **W** matrices of order vxb satisfying the conditions $\mathbf{C}_1^*$, $\mathbf{C}_2^*$ and $\mathbf{C}_3$* simultaneously can be developed by taking the transpose of W'_{ij} matrices.

Remark 4.1: If v ≠ mr, then either (i) (r-1)(v-1) optimal covariates exists for (r-1)(v-1)<(v-1)(b-r-1) or (ii) (v-1)(b-r-1) optimal covariates exists for (r-1)(v-1)≥(v-1)(b-r-1) of RBD with v treatments in b blocks provided $\mathbf{H}_r$ and $\mathbf{H}_{b-r}$ exists.

Remark 4.2: When $\mathbf{H}_r$ do not exist, then (i) $(a-1)^2$ or (a-1)(v-1) optimal covariates exists when $(a-1)^2$ or (a-1)(v-1)<(v-1)(b-r-1) and (ii) (v-1)(b-r-1) optimal covariates exists when $(a-1)^2$ or (a-1)(v-1)≥(v-1)(b-r-1) of RBD with v treatments in b (0 or 2; mod 4) blocks where r can be partitioned in such a way that r = a+e+...+u, provided $\mathbf{H}_a$, $\mathbf{H}_e$,..., $\mathbf{H}_u$ exists and a = min (a, e,..., u).

For easy understanding of the above steps, the following example will be useful.

Example 4.2: Let us consider a RBD with v = 16 and b = 20. When r = 4, the nine **W** matrices are given below:

Step 1. Let us consider a Hadamard matrix of order 16, $\mathbf{H}_{16}$.

$$H_{16}=\left(1,h_1,h_2,\ldots,h_{15}\right)=\begin{pmatrix}
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1\\
1&1&-1&-1&1&1&-1&-1&1&1&-1&-1&1&1&-1&-1\\
1&-1&-1&1&1&-1&-1&1&1&-1&-1&1&1&-1&-1&1\\
1&1&1&1&-1&-1&-1&-1&1&1&1&1&-1&-1&-1&-1\\
1&-1&1&-1&-1&1&-1&1&1&-1&1&-1&-1&1&-1&1\\
1&1&-1&-1&-1&-1&1&1&1&1&-1&-1&-1&-1&1&1\\
1&-1&-1&1&-1&1&1&-1&1&-1&-1&1&-1&1&1&-1\\
1&1&1&1&1&1&1&1&-1&-1&-1&-1&-1&-1&-1&-1\\
1&-1&1&-1&1&-1&1&-1&-1&1&-1&1&-1&1&-1&1\\
1&1&-1&-1&1&1&-1&-1&-1&-1&1&1&-1&-1&1&1\\
1&-1&-1&1&1&-1&-1&1&-1&1&1&-1&-1&1&1&-1\\
1&1&1&1&-1&-1&-1&-1&-1&-1&-1&-1&1&1&1&1\\
1&-1&1&-1&-1&1&-1&1&-1&1&-1&1&1&-1&1&-1\\
1&1&-1&-1&-1&-1&1&1&-1&-1&1&1&1&1&-1&-1\\
1&-1&-1&1&-1&1&1&-1&-1&1&1&-1&1&-1&-1&1
\end{pmatrix}$$

Step 2. Let us construct a Special Array of order 20 from $\mathbf{H}_{16}$ with 4 rows and columns with all zero elements in the middle, $\mathbf{H}_{20}$* i.e., (**1*, $\mathbf{h}_1$*, $\mathbf{h}_2$*,… ,$\mathbf{h}_7$*,0,0,0,0,$\mathbf{h}_8$*,…,$\mathbf{h}_{15}$***).

$$H^*_{20}=\left(1^*,h_1{}^*,h_2{}^*,\ldots,h_7{}^*,0,0,0,0,h_8{}^*,\ldots,h_{15}{}^*\right)$$

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

Step 3. Using $\mathbf{H}_{20}$*and $\mathbf{H}_{16}$, by Kronecker product of these two matrices, we get 15 set of 15 W_{ij}* matrices of order 20x16 where i=1,2,…,15 and j=1,2,…,15. In each of the W_{ij}* matrix there are 4 rows with all elements zero in the middle. First matrix of first set W_{11}* is the following.

$$W^*_{11}=h^*_1\otimes h'_1$$

$$=\begin{pmatrix}1\\-1\\1\\-1\\1\\-1\\1\\-1\\0\\0\\0\\0\\1\\-1\\1\\-1\\1\\-1\\1\\-1\end{pmatrix}\otimes\left(1\ -1\ 1\ -1\ 1\ -1\ 1\ -1\ 1\ -1\ 1\ -1\ 1\ -1\ 1\ -1\right)$$

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

Similarly, we can easily construct the others.

Step 4. As $\mathbf{H}_{16}$ and $\mathbf{H}_4$ both are exists, following the Theorem 3.4.1 (Das *et al.*, 2015), we construct either 9W** matrices of order 4 or 45 W** matrices of order 4x16, which are orthogonal to each other. Here, in this case, 9W** matrices of order 4 has been considered.

$$W_1^{**}=\begin{pmatrix}1&-1&1&-1\\-1&1&-1&1\\1&-1&1&-1\\-1&1&-1&1\end{pmatrix},\quad W_2^{**}=\begin{pmatrix}1&-1&-1&1\\-1&1&1&-1\\-1&1&1&-1\\1&-1&-1&1\end{pmatrix},\quad W_3^{**}=\begin{pmatrix}1&1&-1&-1\\1&1&-1&-1\\-1&-1&1&1\\-1&-1&1&1\end{pmatrix},$$

$$W_4^{**}=\begin{pmatrix}1&-1&1&-1\\1&-1&1&-1\\-1&1&-1&1\\-1&1&-1&1\end{pmatrix},\quad W_5^{**}=\begin{pmatrix}1&-1&-1&1\\-1&1&1&-1\\1&-1&-1&1\\-1&1&1&-1\end{pmatrix},\quad W_6^{**}=\begin{pmatrix}1&1&-1&-1\\-1&-1&1&1\\-1&-1&1&1\\1&1&-1&-1\end{pmatrix},$$

$$W_7^{**}=\begin{pmatrix}1&-1&1&-1\\-1&1&-1&1\\-1&1&-1&1\\1&-1&1&-1\end{pmatrix},\quad W_8^{**}=\begin{pmatrix}1&-1&-1&1\\1&-1&-1&1\\-1&1&1&-1\\-1&1&1&-1\end{pmatrix}\text{ and }\ W_9^{**}=\begin{pmatrix}1&1&-1&-1\\-1&-1&1&1\\1&1&-1&-1\\-1&-1&1&1\end{pmatrix}$$

Step 5. In each W* matrix, insert the first W** matrix of order 4 in the first 4 columns of W_1* matrix and replicate the selected W** matrix 3 times, such that all the 4 rows with all elements zero has been replaced by +1 or -1. Let the resulting matrix be W_{11}'. Repeat the procedure with other W** matrices in the remaining W* matrices till W** matrices or W* matrices has been utilized totally. So, we get 9 W′ matrices of order 20x16 as 9 < 225 which are orthogonal to each other and all the W′ matrix has all column-sums and row-sums equal to zero. Finally, the desired W matrices of order 16x20 satisfying the conditions $\mathbf{C}_1^*$, $\mathbf{C}_2^*$ and $\mathbf{C}_3$* simultaneously can be developed by taking the transpose of W'_{ij} matrices where i=1,2,…,15 and j=1,2,…,15. Here, W_1** matrix is inserted in W^*_{11} matrix and replicate W_1** matrix 3 times, we get the following matrix W'_{11}.

$$W'_{11}=\begin{pmatrix}
1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1\\
-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1\\
1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1\\
-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1\\
1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1\\
-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1\\
1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1\\
-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1\\
1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1\\
-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1\\
1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1\\
-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1\\
1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1\\
-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1\\
1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1\\
-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1\\
1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1\\
-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1\\
1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1\\
-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1
\end{pmatrix}$$

Finally, the desired $\mathbf{W}_1$ matrix of order 16x20 can be developed by taking the transpose $\mathbf{W}'_{11}$ matrix i.e.,$\mathbf{W}_1 = (\mathbf{W}'_{11})'$. Similarly, we can find out the others. Here, we can construct 9**W** matrices. Alternately, we can construct 45**W**matrices by using 45**W**** matrices of order 4x16. For the RBD with v=16 and b=20, the other possible alternatives are shown in the following Table 4.1.

Table 4.1. The other possible alternatives for RBD with v=16 and b=20.

No. of rows (columns) with all elements zero in the SA of order 20 (r)	No. of optimum covariates (c)
8	49 or 105
12	105
16	45
18	15

Corollary 4.2: The optimal covariate design in RBD developed by theorem 4.2 is true for CRD with similar v and b.

Proof: Straight forward from the definition of CRD.

5. CONCLUSION

New global optimal covariate designs in CRD and RBD set-ups have been presented in section 3 and 4. In Theorem 3.1 and Theorem 4.1, the developed designs require only the Hadamard matrices $\mathbf{H}_2$ and $\mathbf{H}_4$. There is no need to existence of Hadamard matrices $\mathbf{H}_v$ and $\mathbf{H}_b$, where v is the number of treatments and b is the number of replications or blocks. The Theorem 4.2 yields several OCDs in RBD set-up by using $\mathbf{H}_v$ and special array of order b; $\mathbf{H}_b$ does not exist. The developed optimal covariate designs based on the above theorems are not available in the existing literature.

ACKNOWLEDGEMENT

The authors are grateful to the referee for giving valuable suggestions about the article which are included in the manuscript.

REFERENCES

Das, K., Mandal, N.K. and Sinha, B.K. (2003). Optimal experimental designs with covariates. *J. Stat. Plan. Inference*, **115**, 273-285.

Das, P., Dutta, G., Mandal, N.K. and Sinha, B.K. (2015). *Optimal covariate design-Theory and applications*. Springer, India.

Dey, A. and Mukherjee, R. (2006). D-optimal designs for covariate models. *Statistics*, **40**, 297-305.

Dutta, G. (2004). Optimum choice of covariates in BIBD set-up. *Cal. Stat. Assoc. Bull.*, **55**, 39-55.

Dutta, G., Das, P. and Mandal, N.K. (2007). Optimum choice of covariates for a series of SBIBDs obtained through projective geometry. *J. Modern Appld. Stat. Methods*, **6**, 649-656.

Dutta, G., Das, P. and Mandal, N.K. (2010b). D-optimal designs for covariate parameters in block design set-up. *Commun. Stat. Theory Methods*, **39**, 3434-3443.

Dutta, G., Das, P. and Mandal, N.K. (2014). D-Optimal designs for covariate models. *Commun. Stat. Theory Methods*, **43**, 165-174.

Haggstrom, G.W. (1975). *Pitfalls Manpow Exp*. RAND Corporation, Santa Monica.

Harville, D.A. (1974). Nearly optimal allocation of experimental units using observed covariate values. *Technometrics*, **16**, 589-599.

Harville, D.A. (1975). Computing optimum designs for covariate models. *IN: Srivastava JN (ed) A survey of statistical design and linear models*. 209-228, Amsterdam, North Holland.

Rao, C.R. (1973). *Linear statistical inference and its applications*. Wiley, New York.

Rao, PSSNVP., Rao, S.B., Saha, G.M. and Sinha, B.K. (2003). Optimal designs for covariates' models and mixed orthogonal arrays. Electronic Notes. *Discret. Math*., **15**, 157-160.

Shah, K.R. and Sinha, B.K. (1989). Theory of optimal designs. *Lecture notes in statistics*, **series 54**, Springer, New York.

Sinha, B.K. (2009). A reflection on the choice of covariates in the planning of experimental designs. *J. Indian Soc. Agric. Stat.*, **63(3)**, 219-225.

Troya, Lopes, J. (1982a). Optimal designs for covariate models. *J. Stat. Plan. Inference*, **6**, 373-419.

Troya, Lopes, J. (1982b). Cyclic designs for a covariate model. *J. Stat. Plan. Inference,* **7**, 49-75.

Wu, CFJ. (1981). Iterative construction of nearly balanced assignments I: Categorical covariates. *Technometrics*, **23**, 37-44.

Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 74(1) 2020 51–56

Estimation of Drug Addicts among Students of Senior Secondary Schools Located in Kumaun Region of Uttarakhand, India Using RRT

Neeraj Tiwari and Tanuj Kumar Pandey
Kumaun University, S.S.J. Campus, Almora

Received 12 March 2019; Revised 09 October 2019; Accepted 04 January 2020

SUMMARY

We applied the randomised response technique (RRT) to estimate the number of drug addicts among the male and female students of senior secondary schools located in Kumaun region of Uttarakhand, India. Warner's RRT was applied to estimate the drug addicts among the senior secondary students in the region. The procedure adopted in the study provides adequate confidentiality to the respondents and reduces survey time. An estimated number of 26% of total students were found to be addicted to drugs with a 95% confidence interval [20%, 32%]. A strong association was found between drug addiction and academic performance of the students. The study is highly useful for planners in state and central governments to assess the gravity of the drug addiction among the school going children of the region and find ways to control the growing menace of drug addiction in the society.

Keywords: Randomized response, Related question model, Estimates, Drug addiction.

1. INTRODUCTION

Drug addiction can be defined as the habitual taking of illegal drugs. Gelder and Cowen (2001) observed that substance abuse or drug abuse results in clinically significant impairment or distress, wherein the person may suffer from tolerance and withdrawal. Foo *et al.* (2012) found that the parents' substance abuse habits were the most influential factor in affecting a child's substance abuse. These researches were based upon direct questioning method of the survey.

Drug addicts in a community can be estimated by direct questioning but some respondents may not answer truthfully or provide false responses regarding their status about drug abuse due to social stigma. To eliminate the evasive bias due to sensitivity of questions, a technique known as randomised response technique (RRT), initially propounded by Warner (1965) is used. In this technique, a randomizing device is used to extract answers of the questions having sensitive nature by protecting the privacy of the respondents. Suppose we are interested in estimating the proportion of individuals in the population who possess a sensitive character A. The population is, therefore, dichotomous, some possessing the character A, and others possessing the complementary character A'. Because of the sensitive and often stigmatic nature of A, direct questions would result in biased estimates of the population proportion(π) as most of the respondents would give untruthful or evasive answers. Though the resulting evasive answer bias is ordinarily difficult to assess, it is potentially removable by allowing the interviewee to maintain privacy through the randomisation device. To eliminate this bias, Warner (1965) suggested a related question model followed by unrelated question model by Greenberg *et al.* (1969), which were subsequently improved by different authors. The nature of the data so obtained by implementing these techniques is either qualitative or quantitative. A summarisation of different randomised response techniques can be seen in Fox and Tracy (1987), Chaudhuri and Mukherjee (1988), and Chaudhuri *et al.* (2016).

The RRT suggests itself as one of the natural choices in case of the sensitive nature of the survey. It protects privacy of the respondent and develops a good

Corresponding author: Neeraj Tiwari
E-mail address: neerajtiwari.amo@gmail.com

rapport with the interviewer. The respondent provides information on a probability basis without revealing their personal status. By allowing the respondent to maintain his privacy, a better cooperation from him as compared to direct questioning is expected. Thus, RRT can be used to assess the severity of drug addiction among the students of senior secondary schools of Kumaun region of Uttarakhand, India. This study also enabled us to check the limitations of RRT in such type of field surveys.

Researchers used RRT in different fields to estimate the proportion of individuals possessing sensitive attribute like drug addiction, sexual abuse, extramarital affair, tax evasion etc. Chow *et al.* (1979) applied a multiple answer model of RRT (known as Hopkins RRT model II) in a rural area in Ethiopia to estimate the incidence of induced abortion among newly married women of childbearing age. Scheers and Dayton (1987) used RRT in estimating academic cheating behavior of university students. Houston and Tran (2001) conducted a survey using RRT for estimating tax evasion. Soudarssanane *et al.* (2003) applied RRT to estimate the prevalence of pre/extra marital sex. Ostapczuk *et al.* (2009) presented a randomized response investigation of the education effect in attitudes towards foreigners. Srivastava *et al.* (2015) used multi-proportions randomized response technique to assess the extent of sexual abuse among children in some districts of Uttar Pradesh state of India. Chhabra *et al.* (2016) used optional unrelated question RRT and asked the question "Have you ever been a victim of sexual abuse by a friend or family member?" to 585 students in a college in Delhi, India. Cobo *et al.* (2016) conducted an RRT survey into stratified sample of 1146 students of Spanish University and asked sensitive quantitative questions about cannabis use. Kirtadze *et al.* (2018) applied RRT to estimate the proportion of alcohol and other drug users in the country of Georgia.

In recent times, several cases concerning drug addiction including some most severe criminal cases related to drug abuse or illegal use of drugs among students of the Kumaun region of Uttarakhand, India have surfaced. Saxena and Upadhyay (2016) reported that the problem of drug addiction is acute in Kumaun region of Uttarakhand, where according to their sources almost 50 per cent of those being admitted in de-addiction centers, are in the age group of 12-19 years. From secondary sources like magazines, newspapers and electronic media, it is revealed that in the community of school going children in India, marijuana has been the most popular drug. These secondary data encouraged us to make an attempt in determining the severity of drug addiction among students of senior secondary schools in Kumaun region of Uttarakhand, India. RRT was used for the study due to the social stigma attached to drug addiction in that age group. In this paper, we apply the Warner's RRT to estimate the number of drug addicts among the students of senior secondary schools located in Kumaun Region of Uttarakhand, India.

2. MATERIAL AND METHODS

The purpose of the present study is to apply Warner's RRT to estimate the proportion of drug addicts in Kumaun region of Uttarakhand. Our approach was slightly different in this case as we constructed an anonymous questionnaire and instructed the respondents in a group about the RRT and ensured them regarding the confidentiality of data. After instructing them, we collected the information in straightforward manner by isolating each respondent from the group. This procedure helped in reducing survey time.

The anonymous questionnaire consisted of 8 questions. In questions numbered from 1-7, demographic and socio-economic information such as gender, type of school, place of living, type of family, family monthly income, academic performance etc. were sought whereas question no. 8 was based on the sensitive issue (drug addiction), which was answered by the respondents using the randomizing device (i.e. by rolling a die). There were two statements each having a binary response (Yes or No) under question No. 8, namely, Statement A and Statement B. If the die rolled 2, 3, 4 or 5, statement A was to be answered by the respondent, which state that "I consume illegal drugs" and if the number surfaced on the die was 1 or 6, statement B was to be answered by the respondent, which was a negation of the statement A, i.e., "I do not consume illegal drugs" (see Appendix-I). Clearly, the probabilities of selection of statement A and statement B are 0.67 and 0.33 respectively. These probabilities under the Warner's model were recommended by Mukhopadhyay (2014) for estimating the sensitive issue.

In this study, the data have been collected in the month of October 2018 by interviewing 200 students from four senior secondary schools (name of schools have been kept confidential due to the sensitive nature of the survey) situated in Nainital and Udham Singh Nagar districts of Kumaun region of Uttarakhand, India. The schools and students from each school were selected randomly using the simple random sampling without replacement technique. Data were collected by distributing the questionnaire personally to the students in a group by instructing them to answer only first 7 questions directly, and for answering the question number 8 respondents were called to the corner of room one by one to use the randomizing device. Before responding to this question, each respondent was briefed about the method and instructed to answer truthfully. Thus, the sensitive question had been answered by respondents with the help of randomizing device, *i.e.* a simple six faced unbiased die. On the basis of the response of question no. 8, collected questionnaires had been categorised into two categories viz. "Yes" Answered questionnaires and "No" Answered questionnaires. Warner's related question model was adopted in the questionnaire and as such the estimate of drug addicts was computed through the Warner's estimate.

2.1 Statistical Approach

Assuming that the respondent answers truthfully through the randomised response device, using Warner's (1965) procedure, the probability of a 'yes' answer is,

$$\lambda = \pi P + (1-\pi)(1-P) = (1-P) + (2P-1)\pi \tag{1}$$

Denoting the number of 'yes' answers in the sample as *r*, an unbiased estimator of λ is,

$$\hat{\lambda} = \frac{r}{n}.$$

Hence, from (1) an unbiased estimator of π is (taking $P \neq ½$),

$$\hat{\pi}_w = \frac{P-1}{2P-1} + \frac{r}{(2P-1)n} \tag{2}$$

When $P = 1$, a direct response survey occurs and we get $\hat{\pi}_w = \frac{r}{n}$, the usual estimate of population proportion π, whose variance is $\pi(1-\pi)/n$. *r*, the number of "Yes" answers in the sample, is a random variable as there has been a probability associated with it. It follows a binomial distribution with parameters (n, λ). The variance of $\hat{\pi}_w$ is given by,

$$Var(\hat{\pi}_w) = \frac{\lambda(1-\lambda)}{n(2P-1)^2}$$

$$= \frac{\pi(1-\pi)}{n} + \frac{1}{n}\left[\frac{1}{16(P-0.5)^2} - \frac{1}{4}\right]$$

$$= \frac{\pi(1-\pi)}{n} + \frac{P(1-P)}{n(2P-1)^2}$$

Since

$$E\left(\frac{\hat{\lambda}(1-\hat{\lambda}}{n-1}\right) = \frac{\lambda(1-\lambda)}{n},$$

an unbiased estimator of $Var(\hat{\pi}_w)$ is,

$$Est.\{Var(\hat{\pi}_w)\} = \frac{\hat{\lambda}(1-\hat{\lambda}}{(n-1)(2P-1)^2}$$

$$= \frac{\hat{\pi}_w(1-\hat{\pi}_w)}{n-1} + \frac{P(1-P)}{(n-1)(2P-1)^2} \tag{3}$$

3. RESULTS

For this field survey, the population consists of the students studying in senior secondary schools in Kumaun region of Uttarakhand. Based on a prior guess at the parameters of interest and on the randomization device parameters being used in collecting the data, Lee *et al.* (2013) recommended minimum sample sizes for the Warner's model. By considering the relevant parameters associated with our randomization device, 200 students were selected by SRSWOR from different Senior Secondary Schools of the region. A regular unbiased die was used as a randomization device for Warner's model with P = 2/3. Table 1 represents the binary response distribution of the sensitive question according to the gender. The estimates of proportion possessing the sensitive attribute and its variance were computed for the model using equations (2) and (3), respectively. The estimates of drug addicts among male and female students are given in Table 2.

As per the estimates developed in the study, 32% male and 18% female students were drug addicts in the region. The difference between proportions of drug addicts among male and female students were significant at 5% level of significance (p-value = 0.03). Thus, it can be concluded that there is a significant difference between the proportion of drug addicts among the male and female students of senior secondary schools in Kumaun Region of Uttarakhand. It has also been estimated that the proportion of total drug addicts among students of senior secondary schools of the region is 26% with a 95% confidence interval [20%, 32%].

Table 1. Binary response distribution of sensitive question according to gender

			Gender		Total
			Male	**Female**	
Response	**No**	**Count**	65	51	116
		% within Gender	56.0%	60.7%	58.0%
	Yes	**Count**	51	33	84
		% within Gender	44.0%	39.3%	42.0%
Total		**Count**	116	84	200
		% within Gender	100.0%	100.0%	100.0%

Table 2. Warner's estimate of drug addicts among male and female students of senior secondary schools in Kumaun Region of Uttarakhand

Gender	P	"Yes" Answers	$\hat{\lambda}$	$\hat{\pi}_w$	$Var(\hat{\pi}_w)$
Male	0.67	51	0.44	0.32	0.0107
Female	0.67	33	0.39	0.18	0.0103
Total	**0.67**	**84**	**0.42**	**0.26**	**0.0106**

In our questionnaire, we have extracted the information regarding the recent academic performance of students in 3 categories. The categories were made according to their recent examination marks percentage (CGPA) and were termed as; Low (less than 50%), Average (50% – 75%) and High (above 75%). On the basis of this information the association between drug addiction and academic performance of students can be tested.

To perform the study, the "Yes" answered questionnaires were classified into 3 categories viz. Low, High and Average, and Chi-square test was applied to the data. Table 3 exhibits the outcome of the analysis. Table 3 reveals that we obtained a p-value lower than the desired level of significance (5%), hence it may be concluded that there exists an association between drug addiction and academic performance of students. Hence, it may be concluded that the drug addiction affects the academic performance of students.

4. CONCLUSIONS AND DISCUSSION

Making normal assumptions regarding the RRT approach and assuming that a few participants made mistakes during the RRT process, the results presented in this study suggest that RRT can be effectively used to estimate the population possessing a sensitive characteristic such as drug addiction. The study suggests that the menace of drug addiction is quite alarming (26%) among the students of senior secondary schools in Kumaun Region of Uttarakhand, India. While 32% of the male students were found to be drug addicts, the female students were also not legging behind with an estimate of 18% drug addicts among them. However, a significant difference was found between the proportion of male and female drug addicts among the senior secondary students in the society. It was also revealed by the study that the drug addiction among students has an effect on their academic performance.

The study is highly useful for planners in state and central governments to assess the gravity of the drug addiction in the schools and find ways to control the growing menace of drug addiction in the society. The state government must take immediate steps to control the situation and the parents and society should also

Table 3. Association between Academic performance and Drug addiction.

			Drug-Addiction		Total	Chi-Square and p-value
			Non- Addicts	**Addicts**		
Academic Performance (AP)	**Low**	**Count**	33	24	57	11.038 **(0.004*)**
		% within AP	57.9%	42.1%	100.0%	
	Average	**Count**	93	24	117	
		% within AP	79.5%	20.5%	100.0%	
	High	**Count**	22	4	26	
		% within AP	84.6%	15.4%	100.0%	
Total		**Count**	148	52	200	
		% within AP	74.0%	26.0%	100.0%	

*Significant at 5% and 1% level of significance.

cooperate with the government initiatives to keep their children away from drugs. The help of non-government organizations working in the area, print and electronic media and social workers may also be taken to educate the children about the harmful effects of drug addiction. Sources of drugs coming to the hands of the children must be identified and strict action should be taken to stop them.

At this stage of initial examination of RRT implemented to a small-scale survey on drug use, the most important finding of the study might be its clear demonstration that an efficient RRT technique can be used in situations where large sample survey has to be conducted for estimating sensitive issues. Since, the RRT is useful in estimating sensitive issue, thus, by implementing such methods we can gather information about other type of sensitive issues already existing or emerging in the state of Uttarakhand as well as in India.

The present study has some limitations as well. Firstly, this study was conducted with a relatively small sample size as compared to the size of the target population, due to insufficient funds and no monetary support from any agency. RRT has provided quite reasonable estimates of drug addicts among students by protecting their privacy in this small-scale research. The research can further be extended to larger scale covering the whole state and the country. In large scale research, the question covering the sensitive issue like drug addiction can be extended/split into many components according to the class of drugs consumed by students. It will help in asserting which illicit drug is more popular among them. Secondly, there are several models developed by various researchers for RRT, however every model has its limitations and drawbacks. Moreover, some of the RRT models have only theoretical framework and lacks the experimental implementation in real life scenarios. In this study, we have applied the basic RRT model suggested by Warner (1965) to make an initial attempt to practically verify the applicability of RRT models in estimating the sensitive characteristics. Consequently, for further studies an efficient technique can be developed and implemented by considering the techniques already in use for a particular field of study. This study has been an attempt to apply RRT for estimating drug addicts in the particular community of students. The RRT approach in this kind of sample survey condition deserves some contemporary improvements and adjustments that will enhance its utility as a reliable and accurate method of estimating drug addicts in the society.

Declaration of interest statement

The authors have no competing interests.

ACKNOWLEDGEMENTS

The authors are grateful to one of the referees for his/her constructive comments, which have led to considerable improvement in presentation of this manuscript.

REFERENCES

Chaudhuri, A. and Mukherjee, R. (1988). *Randomised Response: Theory and Techniques*. North Holland.

Chaudhuri, A., Christofides, T.C. and Rao, C.R. (2016). *Handbook of Statistics, Volume 34, Data Gathering, Analysis and Protection of Privacy Through Randomized Response Techniques: Qualitative and Quantitative Human Traits.* North-Holland.

Chhabra, A., Dass, B.K. and Gupta, S. (2016). Estimating prevalence of sexual abuse by an acquaintance with an optional unrelated question RRT model. *The North Carolina Journal of Mathematics and Statistics*, **2**, pp. 1-9.

Chow, L.P., Gruhn, W. and Chang, W.P. (1979). Feasibility of the randomised response technique in rural Ethiopia. *American Journal of Public Health,* **69**, pp. 273-276.

Cobo, B., Rueda, M.M. and Torrecillas-Lopez, F. (2016). Application of randomised response techniques for investigating cannabis use by Spanish university students. *International Journal of Methods in Psychiatric Research*, doi: 10.1002/mpr.1517.

Foo, Y.C., Tam, C.L., and Lee, T.H. (2012). Family factors and peer influence in drug abuse: a study in rehabilitation centre. *International Journal of Collaborative Research on Internal Medicine and Public Health,* **4(3)**, pp. 190-202.

Fox, J.A. and Tracy, P.E. (1986). *Quantitative Applications in the Social Sciences: Randomized response*. Newbury Park, CA: SAGE Publications, Inc.

Gelder, M. Mayaou, R. and Cowen, P. (2001). *Misuse of Alcohol and Drugs. Shorter Oxford Textbook of Psychiatry*(4th Edition). Oxford University Press, 533-579.

Greenberg, B.G., Abul-Ela, A.L.A., Simmons, W.R. and Horvitz, D.G. (1969). The unrelated question randomized response model: theoretical framework. *J. Amer. Statist. Assoc.,* **64(326)**, pp. 520-539

Houston, J. and Tran, A. (2001). A survey of tax evasion using the randomized response technique. *Advances in Taxation*, **13** pp. 69-94.

Kirtadze, I., Otiashvili, D., Tabatadze, M., Vardanashvili, I., Sturua, L., Zabransky, T. and Anthony, J. C. (2018). Republic of Georgia estimates for prevalence of drug use: Randomized response techniques suggest under-estimation. *Drug and Alcohol Dependence,* **187**, pp. 300-304.

Lee, C.S., Stephen, A.S. and Singh, S. (2013). Simulated minimum sample size requirements in various randomized response models. *Communications in Statistics - Simulation and Computation,* **42(4)**, pp. 771-789.

Mukhopadhyay, P. (2014). *Theory and Methods of Survey Sampling* (2nd Edition). PHI Learning Private Limited, Delhi. ISBN-978-81-203-3676-6.

Oral, E. and Ardah, I. (2015). A two-stage sampling model for the estimation of population proportion and cheating with randomized response and direct questioning. JSM Proceedings, Survey Research Methods Section. Alexandria, *American Statistical Association*, 2598-2605.

Ostapczuk, M., Musch, J. and Moshagen, M. (2009). A randomized-response investigation of the education effect in attitudes towards foreigners. *European Journal of Social Psychology,* **39**, pp. 920-931.

Saxena, S. and Upadhyay, V. (2016, October 28). Drug abuse among youngsters an acute problem in Uttarakhand. *The Times of India.* Retrieved from https://timesofindia.indiatimes.com/.

Scheers, N.J. and Dayton, C.M. (1987). Improved estimation of academic cheating behaviour using the randomized response technique. *Research in Higher Education,* **26**, pp. 61-69.

Soudarssanane, M.B., Naik, B., Sahai, A. and Bazroy, J. (2003). Randomised response technique- An innovative method to measure culturally sensitive variables: results from a pilot survey. *Indian Journal of Community Medicine*, **28(3)**, pp. 138-140.

Srivastava, R., Nigam, A.K. and Singh, N. (2015). Application of randomized response techniques in estimation of prevalence of child sexual abuse. *Statistics and Applications* {ISSN 2454-7395(online)}, **13(1-2)**, pp. 37-45.

Warner, S.L. (1965). Randomized response: a survey technique for eliminating evasive answer bias. *J. Amer. Statist. Assoc.*, **60(309)**, pp. 63-69.

APPENDIX-I

Survey for Estimating Sensitive Issue among Students Questionnaire

Respondent No. ☐

Kindly tick the appropriate boxes below to participate in the survey. Your identity will not be compromised at any stage.

1. **Gender**

   ☐ Male ☐ Female

2. **Type of School**

   ☐ Private Public-School ☐ Government School

3. **Class**

   ☐ 11^{th} ☐ 12^{th}

4. **Place of Living**

   ☐ In City ☐ In Village

5. **Type of Family**

   ☐ Joint Family ☐ Nuclear Family

6. **Family Monthly Income**

   ☐ ₹10,000 and Below ☐ ₹ 11,000 - 30,000 ☐ ₹ 30,000 &Above

7. **What is your academic performance, recently?**

   ☐ Low (40% - 60%) ☐ Average (60% - 70%) ☐ High (Above75%)

8. **Survey Question**

   Roll the dice and please answer Question 1 or Question 2 according to the number you get on the dice provided to you.

If you get 2, 3, 4,or 5, answer **Question 1:** Do you ***consume*** *illegal drugs?*	☐ Yes
If you get 1 or 6, answer **Question 2:** *Do you* ***not consume*** *illegal drugs?*	☐ No

Thank you for the participation…

Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 74(1) 2020 57–62

Dynamics of Nutrient Uptake in Long Term Fertilizer Experiments on Rice In Kerala

V.A. Jesma[1], T.K. Ajitha[1], S Krishnan[1], P.P. Moossa[2] and P Sindhumole[1]

[1]*College of Horticulture, Vellanikkara*
[2]*R.A.R.S., Pattambi*

Received 13 August 2019; Revised 31 December 2019; Accepted 04 January 2020

SUMMARY

The present study is based on the secondary data of grain yield, obtained from AICRP on Long Term Fertiliser Experiment (LTFE) on rice conducted at RARS Pattambi during *kharif* and *rabi* seasons. The objective of the study was to study the influence of plant nutrients namely N, P and K uptake on grain yield of rice using nonlinear regression. Quadratic model was able to capture the relationship between yield and plant nutrients in both the seasons.

Keywords: Plant nutrients, Nonlinear, Regression, LTFE, Quadratic.

1. INTRODUCTION

Increased supply of nutrients has played a key role in enhancing food production to address the necessity of rapidly growing world population. Nutrients exhausted by crops are substituted with chemical fertilizers, to attain nutrient balance and soil fertility. Among various factors that contributeto better yield and quality, the appropriate use of fertilizers is of utmost importance (Sankaran *et al.*, 2005). Determination of optimum levels of NPK fertilizers is crucial for achieving maximum economic gains. According to Ananthi *et al.* (2010) best rate of fertilizer application is that which gives maximum returns at least cost. Among various essential plant nutrients, the macro nutrients N, P and K are crucial for determining the yield and quality. It has been noticed that farmers utilize imbalanced dose of chemical fertilizers which lead to higher insects/disease attack ultimately leading to lower yield (Mannan *et al.*, 2009; Alam *et al.*, 2011). Therefore, there is prodigious need to estimate the best level of NPK fertilizers for maximizing the profit. The first step for this is to estimate the functional relationship existing between the nutrient uptake and crop yield.

The Rothamsted experiments has proved the effectiveness of chemical fertilizers in enhancing the yield of crop plants (Rasmussen *et al.,* 1998; Smil 2002). The long-term experiments at Rothamsted showed that yields were two to three times higher than those without fertilizers or manures (Johnston, 1994).). Also, an increasing supply of nutrient can boost the yield to a threshold value after which the production may be affected in a negative way, *ie.,* the plant doesn't take up all the nutrient that is supplied to them. Nutrient management should ideally provide an input-output balance in long term (Heckman *et al.*, 2003).

Rice, being the staple food of Kerala, the need to increase the yield and quality of rice through sustainable agriculture by proper fertilizer applications and various soil fertility management practices has gained importance. The present study was undertaken to study the influence of plant nutrients *viz;* N, P and K uptake on treatment responses of rice under long term experiments.

Corresponding author: V.A. Jesma
E-mail address: jesmava@gmail.com

2. MATERIALS AND METHODS

The present study was based on secondary data from All India Coordinated Research Project on Long-Term Fertilizer Experiments (AICRP-LTFE) in rice, which was initiated at Regional Agricultural Research Station (RARS), Pattambi in 1997 to study changes in soil quality, crop productivity and sustainability under long term fertilizer experiments in rice. The experiment was carried out in RARS, Pattambi, Kerala using the variety Aiswarya in two planting seasons namely *kharif* and *rabi*. Aiswarya variety of rice developed at RARS, Pattambi is resistant to blast, blight and BPH. It is well suited for first and second crop seasons. The *kharif* season starts from July to October during the south-west monsoon season and the *rabi* cropping season is from October to March (winter).

The following are the details of the experiment:

Number of replications: 4

Number of treatments: 12

Design: Randomized Complete Block Design (RCBD)

Plot size: 125 m^2

Following are the fertiliser treatments:

T_1: 50 percent NPK (as per POP recommendation of KAU)

T_2: 100 percent NPK (90 N: 45 P_2O_5: 45 K_2O)

T_3: 150 percent NPK

T_4: 100 percent NPK + lime @ 600 kg/ha

T_5: 100 percent NPK

T_6: 100 percent NP

T_7: 100 percent N

T_8: 100 percent NPK + FYM @5t/ha to the kharif rice only

T_9: 50 percent NPK + FYM @5t/ha to the kharif rice only

T_{10}: 100 percent NPK + *in situ* growing of *Sesbaniaaculeata,* as green manure crop for kharif rice only

T_{11}: 50 percent NPK + *in situ* growing of *Sesbaniaaculeata,* as green manure crop for kharif rice only

T_{12}: Absolute control

The data recorded on grain yield and nutrient uptake with respect to N,P and K of rice crop in *kharif* and *rabi* seasons for twenty years from 1997- 2017 were collected. Preliminary investigation of the data was done with the help of descriptive and exploratory data analysis. To compare the mean nutrient uptake, analysis of variance was employed. *Post hoc* analysis was carried out using Duncan's Multiple Range Test (DMRT). The relative performance of different treatments with respect to grain yield were compared using independent t test. Nonlinear regression was performed using SPSS software (version 22) to quantify the relative contribution of plant nutrients on crop yield. The regression equation fitted for treatment responses takes the form:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_1^2 + b_5X_2^2 + b_6X_3^2 + b_7X_1X_2 + b_8X_2X_3 + b_9X_1X_3$$

Where b_i, i=1, 2, …, 9 are the partial regression coefficients

X_1, X_2, X_3 are the independent variables under study *viz.,* N uptake, P uptake and K uptake respectively.

3. RESULTS AND DISCUSSION

The descriptive statistics of yield data revealed that the mean grain yield in *kharif* season was 2742.08 kg/ha with a standard deviation of 835.70 kg/ha. In *rabi* season, the mean grain yield was 3077.69 kg/ha with a standard deviation of 371.17 kg/ha.

Table 1. Descriptive statistics of yield data for kharif and rabi Seasons

Statistic	Season	
	Kharif	*Rabi*
Mean	2742.08	3077.69
Standard deviation	835.70	371.17
Skewness	0.83	0.72
Kurtosis	0.79	2.26
CV	30.48	12.06

Exploratory analysis yield data in both the seasons through box plot depicted that treatment responses in *rabi* was higher and more consistent than those in *kharif* season (Fig 1).

After assessing the influence of long-term applications of nutrients on crop yield, it was concluded that the highest yield was obtained under T_8 (100 percent NPK + FYM @5t/ha to the *kharif* rice only) followed by T_{10} (100 percent NPK + *in situ* growing

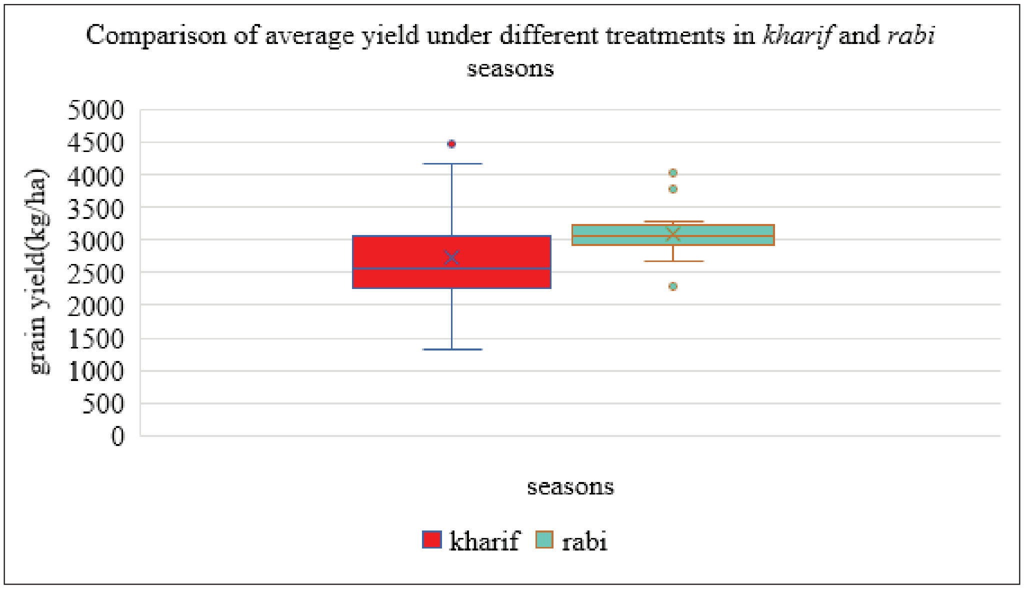


Fig. 1. Comparison of average yield under different treatments in *kharif* and *rabi* seasons

of *Sesbaniaaculeata,* as green manure crop for kharif rice only) in both the seasons. The relative performance of different treatments with respect to grain yield revealed that treatment responses T_7 were significantly different in two seasons. T_7 was reported to be the most imbalanced treatment and even a minute variation in weather affected the yield drastically. When comparing the means for nutrient uptake of N, P and K also, it was established that the highest nutrient uptake of N, P and K was for treatment T_8 followed by T_{10} (Table 2). Yield data recorded over the period 1998-2017 for both *kharif* and *rabi* season clearly validated the superiority of integrated use of FYM and green manuring with chemical fertilizers, which provided greater stability in crop production as compared to 100% NPK. This could be linked with the benefits of organics, which apart from N, P and K supply also improves microbial activities, thereby supplying macro and micro-nutrients such as S, Zn, Cu and B, which are not supplied by inorganic fertilizers.

Simple correlation between the treatment responses and N, P, and K uptake was found to be non-significant, emphasizing the probable curvilinear relationship between these variables and yield. Linear regression between yield and plant nutrients could not account for the variability in yield significantly due to low R^2 values (Table 3). It was observed that treatment T_8 had the maximum uptake of N, P and K when compared to the other treatments showing the significance of the nutrients with respect to yield. So, an attempt was made to quantify the uptake of N, P and K in rice crop.

Table 2. Comparison of mean nutrient uptake of N, P and K in kharif and rabi seasons

Treatments	N uptake		P uptake		K uptake	
	Kharif	*Rabi*	*Kharif*	*Rabi*	*Kharif*	*Rabi*
T1	32.70^{ef}	34.07^{g}	6.95^{f}	7.33^{efg}	56.06^{d}	54.82^{e}
T2	36.51^{cd}	37.61^{ef}	7.33^{def}	7.81^{fg}	63.16^{bc}	59.25^{cd}
T3	36.81^{cd}	38.14^{de}	8.36^{bc}	8.49^{cd}	66.59^{b}	62.63^{c}
T4	36.67^{cd}	38.18^{de}	7.76^{cde}	8.22^{cde}	56.65^{d}	55.60^{de}
T5	34.64^{def}	34.99^{fg}	7.54^{def}	7.47^{fg}	60.59^{cd}	56.00^{de}
T6	38.92^{c}	40.78^{cd}	7.02^{ef}	7.25^{g}	59.02^{cd}	55.91^{de}
T7	32.19^{f}	33.74^{g}	6.84^{f}	7.29^{g}	49.37^{e}	46.98^{f}
T8	51.85^{a}	53.05^{a}	10.78^{a}	11.04^{a}	76.82^{a}	74.08^{a}
T9	39.76^{c}	41.46^{c}	7.99^{cd}	8.55^{c}	60.85^{cd}	59.46^{cd}
T10	46.83^{b}	48.49^{b}	8.86^{b}	9.30^{b}	73.74^{a}	68.48^{b}
T11	35.57^{de}	36.51^{efg}	7.80^{cd}	7.94^{def}	59.82^{cd}	57.47^{de}
T12	25.84^{g}	26.36^{h}	5.10^{g}	5.58^{h}	39.85^{f}	38.55^{g}

Nonlinear regression was used to quantify the relative contribution of the uptake of plant nutrients N, P and K on the treatment responses for both seasons and the results are depicted in Table 4 and Table 5. During *kharif* season, when the quadratic model was fitted for grain yields with respect to different treatments, the R^2 value ranged from 0.67 to 0.89. During *rabi* season, the R^2 values were comparatively higher than that for*kharif* season and ranged from 0.75 to 0.96. This substantiates

Table 3. Linear regression of treatment responses on nutrient uptake of rice in kharif and rabi seasons

Treatments		*Kharif*					*Rabi*				
		b_0	b_1 (N)	b_2 (P)	b_3(K)	R^2	b_0	b_1 (N)	b_2 (P)	b_3(K)	R^2
T_1	Estimates	1718.98	9.31	44.48	2.31	0.14	2292.50	9.72	43.94	0.55	0.08
	Std Error	769.80	17.19	61.20	17.94		476.90	8.89	32.12	11.30	
T_2	Estimates	2222.81	14.04	73.96	-7.81	0.17	2970.31	6.44	33.64	-4.10	0.15
	Std Error	865.95	15.42	75.82	19.00		388.09	7.20	28.17	7.32	
T_3	Estimates	2112.85	11.83	77.06	-5.10	0.21	2917.73	6.45	109.00	-11.80	0.30
	Std Error	929.10	17.38	71.77	17.33		582.16	9.37	45.54	10.19	
T_4	Estimates	2543.67	5.60	63.05	-9.28	0.10	2685.19	11.94	51.54	-7.29	0.18
	Std Error	882.94	17.29	56.75	15.94		427.82	8.78	38.37	8.55	
T_5	Estimates	2258.26	12.02	32.63	-2.00	0.09	2517.15	9.54	7.21	6.44	0.23
	Std Error	1033.72	17.02	61.08	20.10		450.76	8.70	31.13	10.12	
T_6	Estimates	2087.86	12.17	43.62	-5.81	0.13	2385.37	9.35	86.23	-7.09	0.31
	Std Error	647.54	13.02	50.18	9.57		432.50	7.67	40.28	7.13	
T_7	Estimates	2239.57	17.34	18.12	-11.98	0.17	1746.02	20.68	28.72	3.24	0.57
	Std Error	526.20	12.59	54.33	11.80		281.45	8.60	32.29	8.63	
T_8	Estimates	2120.14	3.30	54.77	6.91	0.17	3227.98	11.82	32.65	-4.94	0.20
	Std Error	979.84	16.79	57.25	13.81		521.26	7.36	29.19	7.15	
T_9	Estimates	2775.67	18.91	83.26	-21.13	0.14	3120.88	13.35	4.01	-6.05	0.11
	Std Error	920.16	17.96	75.36	21.81		527.31	9.93	27.75	8.60	
T_{10}	Estimates	2329.65	0.70	80.65	2.05	0.12	3516.09	12.85	28.37	-11.84	0.15
	Std Error	981.16	18.57	79.64	18.91		636.61	9.70	43.96	11.65	
T_{11}	Estimates	2319.85	0.71	64.47	-1.23	0.08	2877.53	13.61	31.00	-6.78	0.15
	Std Error	1095.85	20.54	62.56	18.72		608.16	10.06	42.85	9.38	
T_{12}	Estimates	1407.85	18.77	51.63	-7.14	0.18	1525.37	12.17	18.76	6.19	0.17
	Std Error	602.15	15.51	70.61	14.51		449.42	9.90	30.29	10.12	

Table 4. Model summary of the nonlinear regression of treatment responses on nutrient uptake of rice in kharif season

Treatments		b_0	b_1 (N)	b_2 (P)	b_3(K)	b_4 (N^2)	b_5 (P^2)	b_6 (K^2)	b_7 (NP)	b_8(NK)	b_9(PK)	R^2
T_1	Estimates	-601.99	138.68	-144.96	12.78	-3.94	-45.95	-1.42	8.69	2.16	12.14	0.83
	Std Error	2616.87	68.67	579.55	94.01	1.19	19.61	1.08	6.05	1.53	9.62	
T_2	Estimates	-3816.96	229.49	3.53	40.91	-3.28	-66.41	-1.21	8.92	0.27	14.16	0.84
	Std Error	3649.33	72.52	541.54	97.00	0.84	33.58	1.24	4.41	1.19	14.84	
T_3	Estimates	-971.95	118.73	471.58	-46.04	-3.13	-39.90	-0.35	7.95	1.63	3.33	0.89
	Std Error	2788.21	72.92	328.58	60.44	0.83	20.33	0.52	5.58	0.96	6.54	
T_4	Estimates	-2894.20	166.81	592.71	-37.63	-3.25	-42.99	0.20	12.66	0.60	-1.17	0.81
	Std Error	4311.43	124.20	643.31	107.09	1.68	22.30	1.40	13.04	1.88	8.45	
T_5	Estimates	262.89	124.37	227.38	-63.98	-5.04	-31.03	-1.39	-1.35	5.29	8.59	0.80
	Std Error	4208.33	99.22	1013.64	134.42	1.20	28.96	1.46	8.19	2.07	17.04	
T_6	Estimates	-1930.47	143.05	458.51	-38.99	-2.45	-48.53	-0.25	8.17	0.76	4.86	0.69
	Std Error	4009.35	72.78	894.44	66.45	1.41	18.16	0.41	6.50	2.08	11.62	
T_7	Estimates	831.62	356.34	-1193.98	-60.93	-6.58	-52.32	-2.40	13.50	1.59	35.32	0.70
	Std Error	2012.11	117.56	693.39	79.98	2.02	22.71	1.36	7.71	1.85	12.88	
T_8	Estimates	-4361.47	416.26	-510.11	-51.25	-5.40	-0.20	-1.62	-5.78	3.86	10.61	0.82
	Std Error	3795.35	179.44	541.09	98.00	2.56	35.06	0.74	10.54	2.64	9.91	
T_9	Estimates	-6771.31	73.50	376.60	216.44	-2.92	-75.58	-3.61	13.02	1.87	13.20	0.83
	Std Error	3146.28	77.67	766.49	148.45	1.14	27.53	2.16	5.84	2.20	16.86	
T_{10}	Estimates	-3448.12	162.26	1129.35	-66.42	-3.04	-10.03	0.27	0.92	2.15	-10.61	0.67
	Std Error	3870.68	96.34	884.01	97.28	1.42	24.30	1.62	6.44	2.42	16.21	
T_{11}	Estimates	-2224.33	136.69	-138.80	48.62	-3.43	-72.44	-1.31	25.07	0.01	13.15	0.76
	Std Error	4485.66	138.41	1067.17	108.14	1.66	34.88	0.97	15.32	3.46	17.20	
T_{12}	Estimates	-4557.69	116.41	488.27	144.23	-3.91	-125.56	-2.38	29.67	-0.89	14.12	0.88
	Std Error	2377.05	68.39	417.78	83.34	0.98	29.60	1.16	11.65	1.88	6.01	

Table 5. Model summary of the nonlinear regression of treatment responses on nutrient uptake of rice in rabi season

Treatments		b_0	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	R^2
T_1	Estimates	1295.69	36.12	211.27	-39.55	-1.62	-36.11	-0.48	4.17	1.63	6.59	0.91
	Std Error	2503.54	55.01	344.50	71.55	0.79	11.48	0.52	5.62	0.93	7.33	
T_2	Estimates	-490.34	113.67	159.03	-8.79	-1.76	-28.28	-0.35	4.12	0.66	4.53	0.86
	Std Error	1474.74	66.71	238.53	51.81	0.91	14.25	0.41	5.36	1.12	5.80	
T_3	Estimates	2521.93	-83.12	768.00	-83.57	-0.82	-85.71	-0.14	14.40	1.21	6.93	0.96
	Std Error	1322.14	91.64	166.28	50.13	0.67	15.12	0.40	4.42	0.54	3.62	
T_4	Estimates	2779.93	10.66	271.69	-90.81	0.23	-62.26	-0.03	6.91	-0.35	12.31	0.83
	Std Error	3046.19	111.91	274.82	62.30	1.14	35.65	0.73	9.70	1.28	9.16	
T_5	Estimates	-3444.54	232.11	56.88	67.47	-0.66	-3.81	-0.21	-2.38	-1.94	2.12	0 .75
	Std Error	2195.94	148.25	649.20	71.69	0.75	27.37	1.04	11.25	1.28	12.33	
T_6	Estimates	4075.69	-24.07	-660.93	3.57	-0.26	-76.70	-0.57	25.48	-1.35	16.33	0.92
	Std Error	3180.27	102.46	596.39	46.32	0.73	23.79	0.28	10.22	1.76	8.96	
T_7	Estimates	-579.68	20.63	914.04	-46.75	-5.35	11.16	0.74	11.30	6.12	-28.60	0. 94
	Std Error	563.95	71.45	457.62	35.08	1.56	17.57	0.80	6.05	2.58	16.04	
T_8	Estimates	-3234.09	172.11	17.03	39.36	-1.14	-12.93	-0.34	2.71	-0.33	2.14	0.89
	Std Error	1957.74	105.38	142.64	53.92	0.77	10.31	0.52	3.83	0.59	3.67	
T_9	Estimates	-2983.88	263.02	122.10	-16.78	-2.06	-6.71	0.37	2.54	-0.72	-1.14	0.89
	Std Error	1567.66	95.59	511.55	53.13	1.00	7.52	0.76	5.80	0.95	8.35	
T_{10}	Estimates	-4474.84	139.50	510.05	71.73	-1.77	-11.71	-1.27	-2.22	1.48	-0.72	0.94
	Std Error	1318.92	71.18	317.66	58.02	0.72	24.49	1.06	9.34	1.41	6.36	
T_{11}	Estimates	-2958.63	296.48	-882.27	117.62	-0.91	-16.55	-1.10	6.90	-3.42	15.45	0.85
	Std Error	2469.24	146.49	599.84	65.35	1.32	19.28	0.76	8.09	1.52	6.24	
T_{12}	Estimates	4891.11	-44.71	-3044.50	229.38	3.02	-45.96	-5.81	35.24	-5.89	71.80	0. 76
	Std Error	3196.00	89.36	1517.29	104.78	2.14	14.49	2.97	13.89	2.23	34.88	

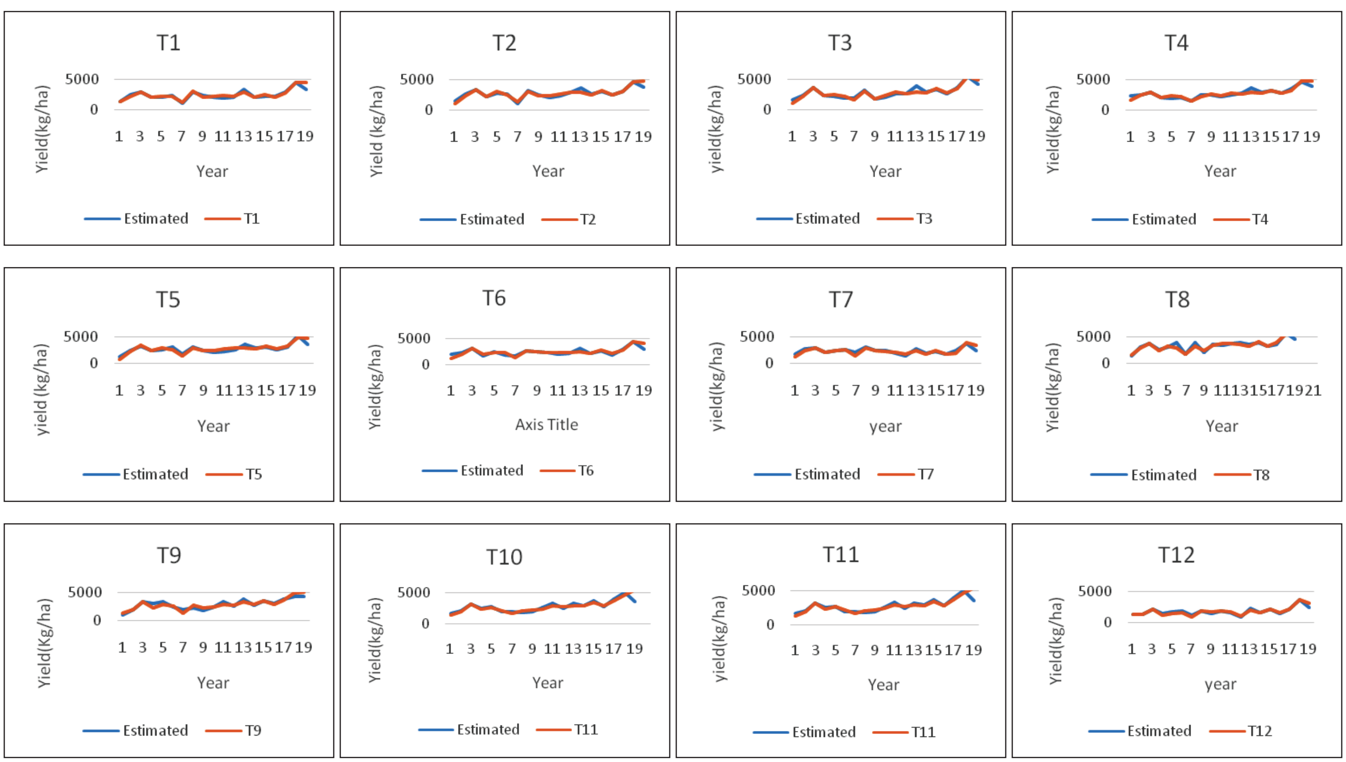


Fig. 2. Comparison of actual and estimated rice yield using quadratic model in *kharif* season

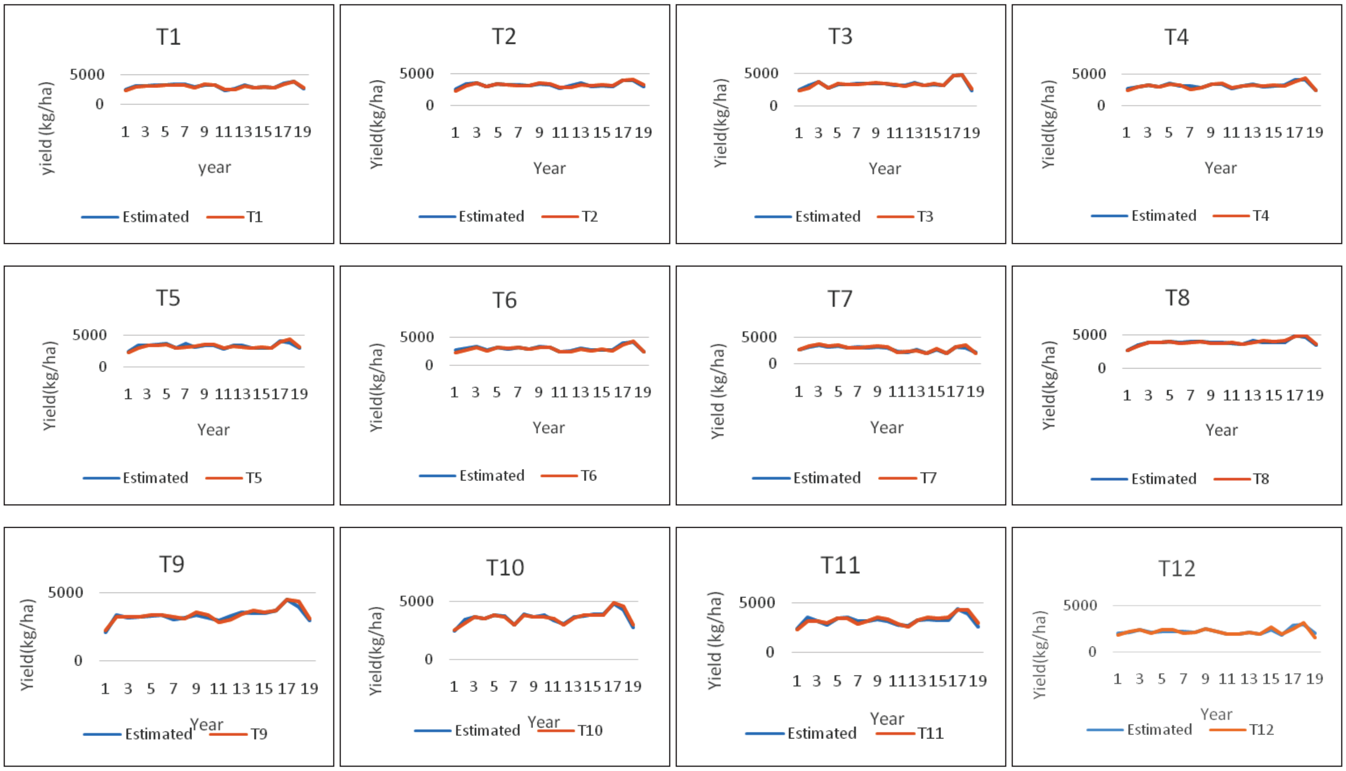


Fig. 3. Comparison of actual and estimated rice yield using quadratic model in *rabi* season

that the relationship existing between crop yield and nutrient uptake is not linear (Fig 2and 3).

4. CONCLUSION

The grain yield in Rabi season was found to be higher and more consistent than that of *kharif* season. The uptake of NPK was found to be maximum under T8 (100 percent NPK + FYM @5t/ha to the *kharif*rice only) which was ranked as the best treatment for maximum grain yield in both the seasons. Despite the idealized vegetation and climatic conditions, the empirical results derived here highlights the nonlinear relationship existing between nutrient uptake of N, P, K and rice yield and can effectively employed to quantify it resulting in high degree of predictability. A similar approach can be encouraged to contribute to our understanding of both regional and larger-scale variations in nutrient uptake dynamics in a more holistic manner.

REFERENCES

Heckman, J.R., Sims, J.T., Beegle, D.B., Coale, F.J., Herbert, S.J., Bruulsema, T.W., Bamka, W.J., 2003. Nutrient removal by corn grain harvest. Agron. J. 95, 87-591.

Rasmussen P.E, Goulding K.W.T, Brown J.R, Grace P.R, Janzen H.H, Korschens M. Long-term agroecosystem experiments: assessing agricultural sustainability and global change. Science. 1998;282:893–896.

Smil V. Nitrogen and food production: proteins for human diets. Ambio. 2002;31:126–131.

Johnston A.E. The Rothamsted classical experiments. In: Leigh R.A, Johnston A.E, editors. Long-term experiments in agricultural and ecological sciences.ch. 2. CAB International; Wallingford, UK: 1994. pp. 9–37

Sankaran N., Meena S, Sakthivel N. 2005. Input management in maize. Mad Agric J 92:464-468.

Mannan M.A., Bhuiya M.S.U., Hossain S.M.A., Akhand M.I.M. 2009. Study on phenology and yielding ability of basmati fine rice genotypes as influenced by planting date in aman season. Ban J Agric Res 34:373- 384.

Alam M.S., Islam M.S., Islam M.A. 2011. Farmers' efficiency enhancement through input management: the issue of USG application in modern rice. Ban J Agric Res 36:129-141.

Ananthi T., Amanullah M.M., Subramanian K.S. 2010. Influence of mycorrhizal and synthetic fertilizers on soil nutrient status and uptake in hybrid maize. Mad Agric J 97:374-378.

Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 74(1) 2020 63–68

An Alternative to Ratio and Product type Estimators of Finite Population Mean in Double Sampling for Stratification

Hilal A. Lone[1], Rajesh Tailor[2] and Med Ram Verma[3]

[1]*Govt. Degree College Sopore, J&K*
[2]*Vikram University Ujjain*
[3]*ICAR-Indian Veterinary Research Institute, Izatnagar*

Received 11 April 2019; Revised 30 July 2019; Accepted 31 January 2020

SUMMARY

In this paper we have proposed an alternative to Ige and Tripathi (1987) estimators with their properties. The expressions for biases and mean squared errors have been obtained upto the first degree of approximation. The proposed estimators have been compared with usual unbiased estimator of population mean in double sampling for stratification and ratio and product type estimators given by Ige and Tripathi (**1987**). To judge the merits of the proposed estimators an empirical study have been carried out.

Keywords: Double sampling for stratification, Bias, Mean squared error.

1. INTRODUCTION

In survey sampling there can be the situations when strata weights are not available or if available, strata weights are outdated and can't be used. This type of situation occurs during the household survey, when investigator does not have information about newly added household in different colonies. This situation leads investigator to use double sampling for stratification. Neyman (1938) developed the theory of double sampling. The problem of estimating finite population mean in double sampling for stratification has been studied by few researchers including Ige and Tripathi (1987),Tripathi and Bahl (1991), Singh and Vishwakarma (2007), Chouhan (2012) ,Sharma (2012), Jatwa (2014), Tailor and Lone (2014a) and Tailor *et al.* (2014b).

Let us consider a finite population $U=\{U_1,U_2,U_3,...U_N\}$ of size N in which strata weight $\frac{N_h}{N},\{h=1,2,3,...L\}$ are unknown. In these conditions we use double sampling for stratification. The procedure for double sampling for stratification is given below

(a) a first phase sample S of size n' using simple random sampling without replacement is drawn and auxiliary variates x and z are observed.

(b) the sample is stratified into L strata on the basis of observed variables x and z. Let n'_h denotes the number of units in h^{th} stratum $(h=1,2,3,...,L)$ such that $n'=\sum_{h=1}^{L} n'_h$.

(c) from each n'_h unit, a sample of size $n_h=v_h n'_h$ is drawn where $0<v_h<1$, $\{h=1,2,3,...,L\}$, is the predetermined probability of selecting a sample of size n_h from each strata of size n'_h and it constitutes a sample S' of size $n=\sum_{h=1}^{L} n_h$. In S' both study variate y and auxiliary variates x and z are observed.

Corresponding author: Med Ram Verma
E-mail address: medramverma@rediffmail.com

Let y be the study variate and x and z are the two auxiliary variate respectively. Let us define

$\bar{x}_{ds} = \sum_{h=1}^{L} w_h \bar{x}_h$: Unbiased estimator of population mean $\bar{X}$ at second phase or double sampling mean of the auxiliary variate x

$\bar{y}_{ds} = \sum_{h=1}^{L} w_h \bar{y}_h$: Unbiased estimator of population mean $\bar{Y}$ at second phase or double sampling mean of the study variate y

$\bar{z}_{ds} = \sum_{h=1}^{L} w_h \bar{z}_h$: Unbiased estimator of population mean $\bar{Z}$ at second phase or double sampling mean of the auxiliary variate z

$\bar{x}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} \bar{x}_{hi}$: Mean of the second phase sample taken from h^{th} stratum for the auxiliary variate x

$\bar{y}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} \bar{y}_{hi}$: Mean of the second phase sample taken from h^{th} stratum for the study variate y

$\bar{z}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} \bar{z}_{hi}$: Mean of the second phase sample taken from h^{th} stratum for the auxiliary variate z

$\bar{X} = \frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h} x_{hi}$: Population mean of the auxiliary variate x

$\bar{Y} = \frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h} y_{hi}$: Population mean of the study variate y

$\bar{Z} = \frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h} z_{hi}$: Population mean of the auxiliary variate z

$\bar{X}_h = \frac{1}{N_h}\sum_{i=1}^{N_h} x_{hi}$: h^{th} stratum population mean for the auxiliary variate x

$\bar{Y}_h = \frac{1}{N_h}\sum_{i=1}^{N_h} y_{hi}$: h^{th} stratum population mean for the study variate y

$\bar{Z}_h = \frac{1}{N_h}\sum_{i=1}^{N_h} x_{hi}$: h^{th} stratum population mean for the auxiliary variate z

$S_x^2 = \frac{1}{N-1}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\left(x_{hi} - \bar{X}_h\right)^2$: Population mean square of the auxiliary variate x

$S_y^2 = \frac{1}{N-1}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\left(y_{hi} - \bar{Y}_h\right)^2$: Population mean square of the study variate y

$S_z^2 = \frac{1}{N-1}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\left(z_{hi} - \bar{Z}_h\right)^2$: Population mean square of the auxiliary variate z

$S_{xh}^2 = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}\left(x_{hi} - \bar{X}_h\right)^2$: h^{th} stratum population mean square of the auxiliary variate x

$S_{yh}^2 = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}\left(y_{hi} - \bar{Y}_h\right)^2$: h^{th} stratum population mean of the study variate y

$S_{zh}^2 = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}\left(z_{hi} - \bar{Z}_h\right)^2$: h^{th} stratum population mean square of the auxiliary variate z

$\rho_{yxh} = \frac{S_{yxh}}{S_{yh}S_{xh}}$: Correlation coefficient between y and x in the stratum h,

$\bar{x}'_h = \frac{1}{n'_h}\sum_{h=1}^{n_h} \bar{x}_{hi}$: First phase sample mean of the h^{th} stratum for the auxiliary variate x

$\bar{z}'_h = \frac{1}{n'_h}\sum_{h=1}^{n_h} \bar{z}_{hi}$: First phase sample mean of the h^{th} stratum for the auxiliary variate z

$f = \frac{n'}{N}$: First phase sampling fraction.

$n = \sum_{h=1}^{L} n_h$: size of the sample S'

$w'_h = \frac{n'_h}{n'}$: h^{th} stratum weight in the first phase sample

$\bar{x}' = \frac{1}{n'_h}\sum_{h=1}^{n_h} w_h \bar{x}'_h$: Unbiased estimator of population mean $\bar{X}$ for the first phase

$\bar{z}' = \frac{1}{n'_h}\sum_{h=1}^{n_h} w_h \bar{z}'_h$: Unbiased estimator of population mean $\bar{Z}$ for the first phase

Ige and Tripathi (1987) defined classical ratio and product estimators in double sampling for stratification as

$$\bar{y}_{Rd} = \bar{y}_{ds}\left(\frac{\bar{x}'}{\bar{x}_{ds}}\right). \tag{1.1}$$

and

$$\bar{y}_{Pd} = \bar{y}_{ds}\left(\frac{\bar{z}_{ds}}{\bar{z}'}\right). \tag{1.2}$$

where z is an auxiliary variate which is negatively correlated with the study variate y and notations $\bar{z}_{ds}$ and $\bar{z}'$ have their usual meanings.

The biases and mean squared errors of estimators $\bar{y}_{Rd}$ and $\bar{y}_{Pd}$ up to the first degree of approximation are given by

$$B\left(\bar{y}_{Rd}\right) = \frac{1}{\bar{X}}\left[\sum_{h=1}^{L}\frac{W_h}{n'}\left(\frac{1}{v_h}-1\right)\left\{R_1 S_{xh}^2 - S_{yxh}\right\}\right], \tag{1.3}$$

$$B\left(\bar{y}_{Pd}\right) = \frac{1}{\bar{Z}}\left[\sum_{h=1}^{L}\frac{W_h}{n'}\left(\frac{1}{v_h}-1\right)S_{yzh}\right], \tag{1.4}$$

$$MSE\left(\bar{y}_{Rd}\right) = S_y^2\left(\frac{1-f}{n'}\right) + \frac{1}{n'}\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)\left[S_{yh}^2 + R_1^2 S_{xh}^2 - 2R_1 S_{yxh}\right], \tag{1.5}$$

and

$$MSE\left(\bar{y}_{Pd}\right) = S_y^2\left(\frac{1-f}{n'}\right) + \frac{1}{n'}\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)\left[S_{yh}^2 + R_2^2 S_{zh}^2 + 2R_2 S_{yzh}\right] \tag{1.6}$$

Srivenkataramana (1980) and Bandhyopadhyaya (1980) used the transformation $x_i^* = \frac{N\bar{X} - nx_i}{N-n}$ and $z_i^* = \frac{N\bar{Z} - nz_i}{N-n}$ on auxiliary variate x and z and obtained dual to classical ratio and product estimator as

, (1.7)

and

$$\hat{\bar{Y}}_P^* = \bar{y}\left(\frac{\bar{Z}}{\bar{z}^*}\right). \tag{1.8}$$

where $\bar{x}^* = \frac{N\bar{X} - n\bar{x}}{N-n}$ and $\bar{z}^* = \frac{N\bar{Z} - n\bar{z}}{N-n}$ are unbiased estimators of population mean $\bar{X}$ and $\bar{Z}$ respectively.

2. PROPOSED ESTIMATORS

Following Srivenkataramana (1980) and Bondyopadhyayh (1980) transformation, we proposed an alternative to Ige and Tripathi (1987) estimators in double sampling for stratification as

$$\bar{y}_{Rd}^* = \bar{y}_{ds}\left(\frac{\bar{x}_{ds}^*}{\bar{x}'}\right)$$

$$\text{or } \bar{y}_{Rd}^* = \frac{\bar{y}_{ds}}{\bar{x}'}\left[\frac{N\bar{x}' - n\bar{x}_{ds}}{N-n}\right] \tag{2.1}$$

and

$$\bar{y}_{Pd}^* = \bar{y}_{ds}\left(\frac{\bar{\mathbf{z}}'}{\bar{z}_{ds}^*}\right)$$

$$\text{or } \bar{y}_{Pd}^* = \frac{\bar{y}_{ds}}{\bar{\mathbf{z}}'^{-1}}\left[\frac{N-n}{N\bar{z}' - n\bar{z}_{ds}}\right] \tag{2.2}$$

Where $\bar{x}_{ds}^* = \frac{N\bar{x}' - n\bar{x}_{ds}}{N-n}$ and $\bar{z}_{ds}^* = \frac{N\bar{z}' - n\bar{z}_{ds}}{N-n}$

To obtain the biases and mean squared errors of the proposed estimators $\bar{y}_{Rd}^*$ and $\bar{y}_{Pd}^*$ we write

$\bar{y}_{ds} = \bar{Y}\left(1+e_o\right)$, $\bar{x}_{ds} = \bar{X}\left(1+e_1\right)$, $\bar{x}' = \bar{X}\left(1+e_1'\right)$, $\bar{z}_{ds} = \bar{Z}\left(1+e_2\right)$ and $\bar{z}' = \bar{Z}\left(1+e_2'\right)$

such that $E\left(e_o\right) = E(e_1) = E\left(e_1'\right) = E(e_2) = E\left(e_2'\right) = 0$ and

$$E\left(e_0^2\right) = \frac{1}{\bar{Y}^2}\left[S_y^2\left(\frac{1-f}{n'}\right) + \frac{1}{n'}\sum_{h=1}^{L}W_h S_{yh}^2\left(\frac{1}{v_h}-1\right)\right],$$

$$E\left(e_1^2\right) = \frac{1}{\bar{X}^2}\left[S_x^2\left(\frac{1-f}{n'}\right) + \frac{1}{n'}\sum_{h=1}^{L}W_h S_{xh}^2\left(\frac{1}{v_h}-1\right)\right],$$

$$E\left(e_2^2\right) = \frac{1}{\bar{Z}^2}\left[S_z^2\left(\frac{1-f}{n'}\right) + \frac{1}{n'}\sum_{h=1}^{L}W_h S_{zh}^2\left(\frac{1}{v_h}-1\right)\right],$$

$$E\left(e_0 e_1\right) = \frac{1}{\bar{Y}\bar{X}}\left[\left(\frac{1-f}{n'}\right)S_{yx} + \frac{1}{n'}\sum_{h=1}^{L}W_h S_{yxh}\left(\frac{1}{v_h}-1\right)\right],$$

$$E\left(e_0 e_2\right) = \frac{1}{\bar{Y}\bar{Z}}\left[\left(\frac{1-f}{n'}\right)S_{yz} + \frac{1}{n'}\sum_{h=1}^{L}W_h S_{yzh}\left(\frac{1}{v_h}-1\right)\right],$$

$$E\left(e_1 e_2\right) = \frac{1}{\bar{X}\bar{Z}}\left[\left(\frac{1-f}{n'}\right)S_{xz} + \frac{1}{n'}\sum_{h=1}^{L}W_h S_{xzh}\left(\frac{1}{v_h}-1\right)\right],$$

$$E\left(e_0 e_1'\right) = \frac{1}{\bar{Y}\bar{X}}\left(\frac{1-f}{n'}\right)S_{yx}, \quad E\left(e_1'^2\right) = \frac{1}{\bar{X}^2}S_x^2\left(\frac{1-f}{n'}\right),$$

$$E\left(e_2'^2\right)=\frac{1}{\bar{Z}^2}S_z^2\left(\frac{1-f}{n'}\right),\ E\left(e_1e_1'\right)=\frac{1}{\bar{X}^2}\left(\frac{1-f}{n'}\right)S_x^2,$$

$$E\left(e_2e_2'\right)=\frac{1}{\bar{Z}^2}S_z^2\left(\frac{1-f}{n'}\right),\ E\left(e_1'e_2'\right)=\frac{1}{\bar{X}\bar{Z}}\left(\frac{1-f}{n'}\right)S_{xz},$$

$$E\left(e_0e_2'\right)=\frac{1}{\bar{Y}\bar{Z}}\left(\frac{1-f}{n'}\right)S_{yz}\ \text{and}\ E\left(e_1e_2'\right)=\frac{1}{\bar{X}\bar{Z}}\left(\frac{1-f}{n'}\right)S_{xz}.$$

The biases and mean squared errors of the proposed estimators $\bar{y}_{Rd}^*$ and $\bar{y}_{Pd}^*$ upto the first degree of approximation are obtained as

$$B\left(\bar{y}_{Rd}^*\right)=-\frac{g}{\bar{X}}\frac{1}{n'}\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)S_{yxh}, \tag{2.3}$$

$$B\left(\bar{y}_{Pd}^*\right)=\frac{1}{\bar{Z}}\frac{1}{n'}\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)\left[g^2R_2\,S_{zh}^2+g\,S_{yzh}\right], \tag{2.4}$$

$$MSE\left(\bar{y}_{Rd}^*\right)=S_y^2\left(\frac{1-f}{n'}\right)+\frac{1}{n'}\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)\left[S_{yh}^2+g^2R_1^2S_{xh}^2-2gR_1S_{yxh}\right], \tag{2.5}$$

and

$$MSE\left(\bar{y}_{Pd}^*\right)=S_y^2\left(\frac{1-f}{n'}\right)+\frac{1}{n'}\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)\left[S_{yh}^2+g^2R_2^2S_{zh}^2+2gR_2S_{yzh}\right]. \tag{2.6}$$

3. EFFICIENCY COMPARISONS

The variance of usual unbiased estimator $\bar{y}_{ds}$ in double sampling for stratification is given as

$$V\left(\bar{y}_{ds}\right)=S_y^2\left(\frac{1-f}{n'}\right)+\frac{1}{n'}\sum_{h=1}^{L}W_h\,S_{yh}^2\left(\frac{1}{v_h}-1\right). \tag{3.1}$$

Efficiency comparisons of proposed dual to ratio estimator $\bar{y}_{Rd}^*$

Comparisons of (2.5) with equation (1.5) and (3.1) shows that

(i) $MSE\left(\bar{y}_{Rd}^*\right)<V\left(\bar{y}_{ds}\right)$ if

$$S_y^2\left(\frac{1-f}{n'}\right)+\frac{1}{n'}\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)\left[S_{yh}^2+g^2R_1^2S_{xh}^2-2gR_1S_{yxh}\right]<S_y^2\left(\frac{1-f}{n'}\right)+\frac{1}{n'}\sum_{h=1}^{L}W_h\,S_{yh}^2\left(\frac{1}{v_h}-1\right)$$

$$\Rightarrow\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)\left[S_{yh}^2+g^2R_1^2S_{xh}^2-2gR_1S_{yxh}\right]<\sum_{h=1}^{L}W_h\,S_{yh}^2\left(\frac{1}{v_h}-1\right)$$

$$\Rightarrow R_1g\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)S_{xh}^2<2\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)S_{yxh} \tag{3.2}$$

(ii) $MSE\left(\bar{y}_{Rd}^*\right)<MSE\left(\bar{y}_{Rd}\right)$ if

$$S_y^2\left(\frac{1-f}{n'}\right)+\frac{1}{n'}\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)\left[S_{yh}^2+g^2R_1^2S_{xh}^2-2gR_1S_{yxh}\right]<S_y^2\left(\frac{1-f}{n'}\right)+\frac{1}{n'}\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)\left[S_{yh}^2+R_1^2S_{xh}^2-2R_1S_{yxh}\right]$$

$$\Rightarrow\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)\left[S_{yh}^2+g^2R_1^2S_{xh}^2-2gR_1S_{yxh}\right]<\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)\left[S_{yh}^2+R_1^2S_{xh}^2-2R_1S_{yxh}\right]$$

$$\Rightarrow\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)\left[g^2R_1^2S_{xh}^2-2gR_1S_{yxh}\right]<\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)\left[R_1^2S_{xh}^2-2R_1S_{yxh}\right]$$

$$\Rightarrow R_1(g^2-1)\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)S_{xh}^2<2(g-1)\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)S_{yxh} \tag{3.3}$$

Efficiency comparison of proposed dual to product estimator $\bar{y}_{Pd}^*$

Comparisons of equations (2.6) with equations (1.6) and (3.1) shows that

(i) $MSE\left(\bar{y}_{Pd}^*\right)<V\left(\bar{y}_{ds}\right)$ if

$$S_y^2\left(\frac{1-f}{n'}\right)+\frac{1}{n'}\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)\left[S_{yh}^2+g^2R_2^2S_{zh}^2+2gR_2S_{yzh}\right]<S_y^2\left(\frac{1-f}{n'}\right)+\frac{1}{n'}\sum_{h=1}^{L}W_h\,S_{yh}^2\left(\frac{1}{v_h}-1\right)$$

$$\Rightarrow\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)\left[S_{yh}^2+g^2R_2^2S_{zh}^2+2gR_2S_{yzh}\right]<\sum_{h=1}^{L}W_h\,S_{yh}^2\left(\frac{1}{v_h}-1\right)$$

$$\Rightarrow\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)\left[g^2R_2^2S_{zh}^2+2gR_2S_{yzh}\right]<0$$

$$\Rightarrow R_2g\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)S_{zh}^2<-2\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)S_{yzh} \tag{3.4}$$

(ii) $MSE\left(\overline{y}_{Pd}^{*}\right) < MSE\left(\overline{y}_{Pd}\right)$ if

$$S_y^2\left(\frac{1-f}{n'}\right)+\frac{1}{n'}\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)\left[S_{yh}^2+g^2R_2^2S_{zh}^2+2gR_2S_{yzh}\right]<$$

$$S_y^2\left(\frac{1-f}{n'}\right)+\frac{1}{n'}\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)\left[S_{yh}^2+R_2^2S_{zh}^2+2R_2S_{yzh}\right]$$

$$\Rightarrow\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)\left[S_{yh}^2+g^2R_2^2S_{zh}^2+2gR_2S_{yzh}\right]<$$

$$\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)\left[S_{yh}^2+R_2^2S_{zh}^2+2R_2S_{yzh}\right]$$

$$\Rightarrow\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)\left[g^2R_2^2S_{zh}^2+2gR_2S_{yzh}\right]<$$

$$\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)\left[R_2^2S_{zh}^2+2R_2S_{yzh}\right]$$

$$\Rightarrow R_2(g^2-1)\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)S_{zh}^2<-2(g-1)\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)S_{yzh} \tag{3.5}$$

where $R_1=\frac{\overline{Y}}{\overline{X}}$, $R_2=\frac{\overline{Y}}{\overline{Z}}$ and $g=\frac{n}{N-n}$.

4. EMPIRICAL STUDY

To exhibit the performance of the proposed estimators in comparison to other considered estimators, two population data sets are being used. The descriptions of population are given below.

Population I- [Source: Tailor *et al.* (2014b)]

y: Production (MT/hectare), x: Production in ‘000Tons and z: Area in ‘000hectare

Constant	Stratum I	Stratum II
n_h	4	4
n'_h	7	7
N_h	10	10
$\overline{Y}_h$	1925.8	3115.6
$\overline{X}_h$	214.4	333.8
$\overline{Z}_h$	51.80	60.60
S_{yh}	615.92	340.38
S_{xh}	74.87	66.35
S_{zh}	0.75	4.84
S_{yxh}	39360.68	22356.50
S_{yzh}	411.16	1536.24
S_{xzh}	38.08	287.92
ρ_{yxh}	0.85	0.98
ρ_{yzh}	0.89	0.93
S_y^2	668351.00	

Population- II [Chouhan, S. (2012)

y: Snowy days,

x: rainy days and

z: Total annual sunshine hours

Constant	Stratum I	Stratum II
n_h	4	4
n'_h	7	7
N_h	10	10
$\overline{Y}_h$	142.80	102.60
$\overline{X}_h$	149.70	91.00
$\overline{Z}_h$	1630.00	2036.00
S_{yh}	6.09	12.60
S_{xh}	13.46	6.57
S_{zh}	102.17	103.46
S_{yxh}	18.44	23.30
S_{yzh}	-239.30	-655.30
S_{xzh}	-1073.00	-240.30
ρ_{yxh}	0.22	0.28
ρ_{yzh}	-0.38	-0.50
S_y^2	528.43	

Table 1 reveals that the proposed ratio estimator $\overline{y}_{Rd}^{*}$ has maximum percent relative efficiency in comparison to usual unbiased estimator $\overline{y}_{ds}$ and Ige and Tripathi (1987) ratio estimator $\overline{y}_{Rd}$ for populations 1. Proposed product type estimator $\overline{y}_{Pd}^{*}$ also has highest percent relative efficiency in comparison to usual unbiased estimator $\overline{y}_{ds}$ and Ige and Tripathi (1987) product estimator $\overline{y}_{Pd}$.

Table 1. Percent relative Efficiencies of $\bar{y}_{ds}$, $\bar{y}_{Rd}$, $\bar{y}_{Pd}$, $\bar{y}^*_{Rd}$ and $\bar{y}^*_{Pd}$ with respect to $\bar{y}_{ds}$

Estimators	$\bar{y}_{ds}$	$\bar{y}_{Rd}$	$\bar{y}_{Pd}$	$\bar{y}^*_{Rd}$	$\bar{y}^*_{Pd}$
Population I	100.00	138.99	82.20	**158.12**	*
Population II	100.00	80.66	104.24	*	**106.66**

*** Not applicable**

Table 2. Empirical exhibition of theoretical conditions given in Section 3

Conditions for proposed dual to ratio estimator $\bar{y}^*_{Rd}$	**Population- I**
$MSE\left(\bar{y}^*_{Rd}\right) < V\left(\bar{y}_{ds}\right)$ if $R_1 g \sum_{h=1}^{L} W_h \left(\frac{1}{v_h} - 1\right) S^2_{xh} < 2 \sum_{h=1}^{L} W_h \left(\frac{1}{v_h} - 1\right) S_{yxh}$	**23002.4<46287.9**
$MSE\left(\bar{y}^*_{Rd}\right) < MSE\left(\bar{y}_{Rd}\right)$ if $R_1 (g^2 - 1) \sum_{h=1}^{L} W_h \left(\frac{1}{v_h} - 1\right) S^2_{xh} < 2(g-1) \sum_{h=1}^{L} W_h \left(\frac{1}{v_h} - 1\right) S_{yxh}$	**-19169.7<-15429.3**
Conditions for proposed dual to product estimator $\bar{y}^*_{Pd}$	**Population- II**
$MSE\left(\bar{y}^*_{Pd}\right) < V\left(\bar{y}_{ds}\right)$ if $R_2 g \sum_{h=1}^{L} W_h \left(\frac{1}{v_h} - 1\right) S^2_{zh} < -2 \sum_{h=1}^{L} W_h \left(\frac{1}{v_h} - 1\right) S_{yzh}$	**359.94<670.92**
$MSE\left(\bar{y}^*_{Pd}\right) < MSE\left(\bar{y}_{Pd}\right)$ if $R_2 (g^2 - 1) \sum_{h=1}^{L} W_h \left(\frac{1}{v_h} - 1\right) S^2_{zh} < -2(g-1) \sum_{h=1}^{L} W_h \left(\frac{1}{v_h} - 1\right) S_{yzh}$	**-294.3<-223.69**

5. CONCLUSION

We have proposed an alternative to Ige and Tripathi (1987) estimators with their properties. In Section 3 the theoretical efficiency comparisons of the proposed estimators with other considered estimators have been given. The conditions under which the proposed estimators have less mean squared errors in comparison to usual unbiased estimator and Ige and Tripathi (1987) ratio and product type estimators are calculated empirically and tabulated in Table 2. The proposed product type estimator $\bar{y}^*_{Pd}$ also has highest percent relative efficiency in comparison to usual unbiased estimator $\bar{y}_{ds}$ and Ige and Tripathi (1987) product estimator $\bar{y}_{Pd}$. Thus the proposed estimators are recommended for use in practice for estimating the finite population mean provided the conditions given in section 3 are satisfied.

ACKNOWLEDGMENT

The authors are highly thankful to the learned referees and the Editor for their valuable on the original version of the paper.

REFERENCES

Bandyopadhyay, S. (1980). Improved ratio and product estimators, Sankhya Series C, **42(2)**, 45-49.

Chouhan, S. (2012). Improved estimation of parameters using auxiliary information in sample surveys. Ph.D. Thesis, Vikram University, Ujjain, M.P. India.

Ige, A.F. and Tripathi, T.P. (1987). On double sampling for stratification and use of auxiliary information. J. Ind. Soc. Agricul. Stat., **39**, 191-201.

Jatwa, N.K. (2014). Estimation of population parameters in presence of auxiliary information. Ph.D. Thesis, Vikram University, Ujjain, M.P. India.

Neyman, J. (1938). Contribution to the theory of sampling human population. J. Amer. Statist. Assoc., **33**, 101-116.

Srivenkataramana, T. (1980). A dual to ratio estimator in sample surveys. Biomet. J. **67(1)**, 199-204.

Sharma, B.K. (2012).Use of auxiliary information for estimation of some population parameters in sample surveys. Ph.D. Thesis, Vikram University, Ujjain, M.P. India.

Singh, H.P. and Vishwakarma, G.K. (2007). A general procedure for estimating the mean using double sampling for stratification. Model Assist. Statist. & Appl., **2(4)**, 225-237.

Tailor R. and Lone, H. A. (2014a**)**. Ratio –cum- product estimator of finite population mean in double sampling for stratification. J. of Rel. & Statist. Stds., **7(1)**, 93-101.

Tailor, R., Chouhan, S. and Kim, J. M. (2014b). Ratio and product type exponentials estimators of population mean in double sampling for stratification. Comm. for Statist. Appl. & Meths. **21(1)**, 1-9.

Tripathi, T.P. and Bahl, S. (1991). Estimation of mean using double sampling for stratification and multivariate auxiliary information. Comm. Statist. Theo. & Meth., **20**, 2589-2602.

Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 74(1) 2020 69–76

Non-Parametric Stability Approach for Horticultural Crop Varietal Release

R. Venugopalan, M. Pitchaimuthu and M. Chaithra
ICAR-Indian Institute of Horticultural Research, Bengaluru

Received 16 October 2019; Revised 23 January 2020; Accepted 31 January 2020

SUMMARY

An attempt has been made to propose a non-parametric stability index for crop varietal release based on the performance of a line for multiple traits coupled with stability. Efficacy of this index has been demonstrated with real time data. It is evident from the results that the rank based non-parametric measures computed based on the relative performance of a genotype as compared to others, may be more practically meaningful to come out with stable lines either for release as variety or as a promising line in the ensuing crop hybridization trails. Measure of importance of the traits worked out may also serve as a selection criterion for the breeders in their future hybridization trails. It is suggested to make use of this method in varietal release program and can be extended for Multi-location trail (MLT) based release of crop varieties.

Keywords: Non-parametric stability, Okra, R- codes, Ranks, Varietal release.

1. INTRODUCTION

Horticultural crop improvement research is mainly aimed to exploit the genetic diversity available in the germplasm by employing various biometrical analysis techniques and culminate with identifying stable lines for release as variety either at institute level or across locations. In doing so, the presence of genotype X environment (GXE) interaction makes it difficult to assess the genetic potential of a variety. Due to this, it may so happen that a particular line may be high yielding but may lack in quality and other important crop protection traits (at least to the bench mark values of several traits, as set by the check variety, upon which improvement being attempted). Further, breeders too may be interested to suggest the farmers, a line which performs consistently well in all the evaluated traits over all the years/seasons/locations, including the trait(s) for which improvement was attempted, instead of recommending a line which performs only in few traits. This calls for employing comprehensive stability analysis in crop improvement research.

The conventional parametric approach of stability analysis is based on various stability measures developed since 1966 and used extensively in various horticultural crop improvement research (Onion: Venugopalan and Veere Gowda (2005); Watermelon (Venugopalan and Pitchaimuthu (2009); Chilli: Venugopalan and Madhavi Reddy (2010)). It may be noticed that through this parametric approach contribution of each genotype to GXE interaction was assessed solely based on their performance and stability over years, and most importantly, for each trait individually. However breeders are interested in assessing the contribution of each genotype to GE interaction based on their relative performance coupled with stability over years and to give recommendation based on collective performance across traits. This calls for employing suitable non-parametric method. Accordingly in this communication, by discussing various non-parametric methods, a new index is proposed with case studies in real time experiments carried out in Okra at ICAR-IIHR, Bengaluru, which

Corresponding author: R. Venugopalan
E-mail address: venugopalan.r@icar.gov.in

could be potentially used in any crop varietal release. R-codes were built up for ease of analysis.

2. MATERIALS AND METHODS

Non-Parametric approach of stability analysis:

A number of nonparametric measures for assessing yield stability have been proposed (Thennarasu, 1995; Nassar and Huhn, 1987). These statistical measures are based on the ranks of the genotypes in each environment tested. The ranking is based on values of Y_{ij} with lowest Y_{ij} value receiving the rank 1, the next higher value 2 and so on. The nonparametric measures based on yield ranks of the genotypes in each environment are worked out are below:

$$NP_i^{(1)} = \frac{1}{n}\sum_{j=1}^{n}\left|r_{ij}{}^{*} - Md_i{}^{*}\right|$$

$$NP_i^{(2)} = \frac{1}{n}\left[\sum_{j=1}^{n}\left|r_{ij}^{*} - Md_i^{*}\right| \Big/ Md_i\right]$$

$$NP_i^{(3)} = \frac{\sqrt{\Sigma\left(r_{ij}^{*}-\bar{r}_{i.}^{*}\right)^2/n}}{\bar{r}_{i.}}$$

$$NP_i^{(4)} = \frac{2}{n(n-1)}\left[\sum_{j=1}^{n-1}\sum_{(j'=j+1}^{n}\left|r_{ij}^{*} - r_{ij'}^{*}\right|/\bar{r}_{i.}\right]$$

$$S_i^{(1)} = 2\sum_{j}^{n-1}\sum_{j'=j+1}^{n}\left|r_{ij} - r_{ij'}\right|/[n(n-1)]$$

$$S_i^{(2)} = \sum_{j=1}^{n}\left(r_{ij} - \bar{r}_{i.}\right)^2/(n-1)$$

$$S_i^{(3)} = \sum_{j=1}^{n}\left(r_{ij} - \bar{r}_{i.}\right)^2/\bar{r}_{i.}$$

$$S_i^{(6)} = \frac{\sum_{j=1}^{n}\left|r_{ij} - \bar{r}_{i.}\right|}{\bar{r}_{i.}}$$

The rank r_{ij} is determined based on the rank of i^{th} genotype in j^{th} environment (Y_{ij}). The uncorrected Y_{ij} has the drawback that they may show significance even when there is no GE interaction. Hence, rank r_{ij*} is determined based on corrected phenotypic values $Y^*_{ij}=[Y_{ij}-\bar{Y}_{i.}]$, $\bar{Y}_{i.}$ being the mean performance of i^{th} genotype. The corrected values depend only on the GE interaction and error components. Mdi* is the median ranks for adjusted values. These measures are widely used to assess the stability for different characters individually in crop improvement research. A detailed study from practical point of view is discussed in Ravi *et al.,* 2013.

Pros and cons of Non-Parametric approach of stability analysis:

There is an ample justification for the use of non-parametric measures in the assessment of yield stability of crop varieties. Chief advantages are: (i) No assumptions about the phenotypic observations are needed, (ii) Sensitivity to measurement errors or to outliers is much less compared to parametric measures, (iii) Additions or deletions of one or a few genotypes do not cause distortions to non-parametric measures. (iv) Most of the time, the breeder, is concerned with crossover interaction, an estimate of stability based on rank-information, therefore, seems more relevant, (v) These measures are particularly useful in situations where parametric measures fail due to the presence of large non-linear GEI. For these reasons, non-parametric measures are widely employed in the selection of crop varieties especially when the interest lies in cross over interaction (Raiger and Prabhakaran, 2001). It is a known fact that the non-parametric methods are less powerful than their parametric counterparts. Simulation studies conducted against this background by Raiger and Prabhakaran (2001) have shown that when the number of genotypes in the trial is fairly large, the power efficiency of the nonparametric measures will be quite close to those of the parametric measures.

Non-Parametric approach for crop varietal release developed at ICAR-IIHR

In the foregoing non-parametric approaches discussed for crop stability analysis, it may be pertinent to observe that these statistical measures are based on the ranks of the genotypes in each environment tested, either deduced from the average rank or median rank. Further, all these measures are computed individually for all the traits based on the rank performance of each genotype. However, it is obvious for any researchers to attach more weight a group of traits, which were lacking in the released varieties, as compared to other traits. Further, arbitrarily assigning weights to the evaluated traits may favour the researchers in the final recommendation. Also, from practical point of view, crop breeders may be interested to suggest the farmers, a line which performs consistently well in all the evaluated traits over all the years/seasons/ locations, instead of a line which performs only in few traits. Hence, by taking into these considerations, positive and negative traits, an approach was adopted wherein based on the stability over replications in a year/location coupled with consistency over years/ locations, suitable weights were worked for the traits. To this end, an attempt has been made to suggest a non-parametric based index (termed as Venugopalan index) by assessing the contribution of each genotype to GE interaction based on their relative performance (performance of a genotype compared to others) and stability over years, simultaneously based on various traits in Okra crop improvement research. The step-by step procedure is described as a flow chart.

Objective: Selection of best line over different traits across different environments (years).

Data requirement:

- Data of minimum 3 years/seasons of a location (or over location) with 3 replications each for all traits.
- Pre-defined objective of the data (to be decided based on the objective of the research envisaged by the breeder) to decide the positive or negative traits among the evaluated traits to be studied. For example, disease/pest incidence trait/days to flowering, dwarf cultivar (if aimed at), should take reverse ranking (negative trait) as compared to yield, fruit weight, plant spread, no of nodes. This has to be solely decided by the crop breeders. However, weightage of the traits would be decided based on the approach envisaged as below.

Okra: Eight hybrids of okra were evaluated over three continuous periods 2014-15,2015-16 and 2016-17 for eight different traits *viz.,* Days to 1st flowering, Fruit Length (cm), Fruit Diameter (cm); No of branches, Plant height (cm); Fruit weight (g); Yield (t/h); Incidences of Yellow Vein Mosaic Virus, YVMV (%) at the experimental plot of Division of Vegetable Crops, ICAR-IIHR was considered for this present study.

Steps

- Standardization of data: It is required as characters are measured in different scales.

$$Z = \frac{(x - \bar{x})}{\sigma}$$

- GLM without interaction: Run univariate ANOVA for all characters by taking different environments as replication (average of replication in every year is pre-considered).
- Precision factor: Take $y_i = \frac{1}{\sqrt{MSE}}$ so that trait with least error will get highest importance.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

- Weightage: Proportion of individual y_i is taken over total Y (given as pie chart)

$Y = \sum yi$, $w_i = \frac{yi}{Y}$ x 100

- Difference: Take the difference of individual value ($\bar{x}_i$) and the check ($\bar{x}_c$)

Positive character: Individual value - check value, $d = \bar{x}_i - \bar{x}_c$

Negative character: Check value - Individual value, $d = \bar{x}_c - \bar{x}_i$

- Superiority %: This is calculated by dividing the differenced value by check value and multiplying by 100, $S = \frac{d}{\bar{x}_c}$ x 100

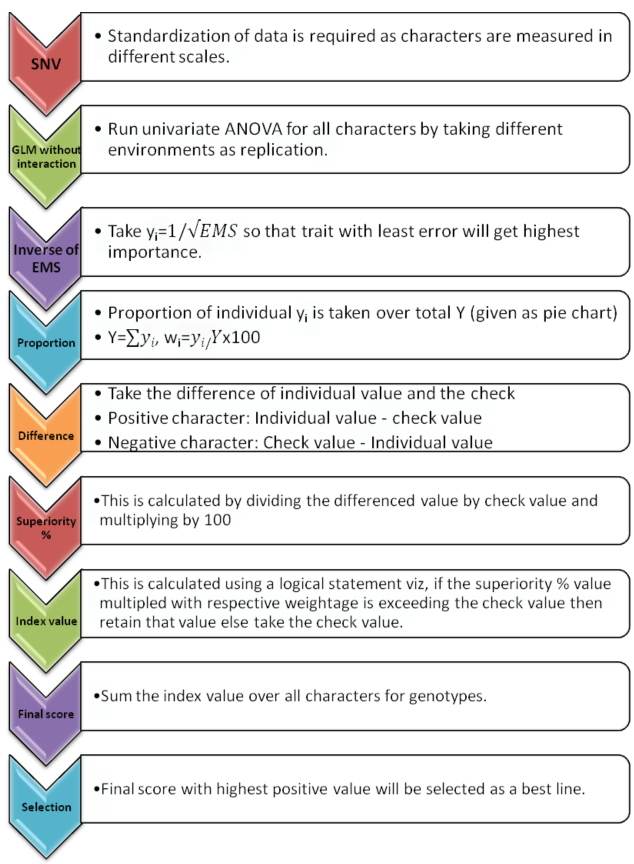


Fig. 1. Flow chart of Non-Parametric approach for crop varietal release developed at ICAR-IIHR

- Index value: This is calculated using a logical statement viz, if the superiority % value multiplied with respective weightage is exceeding the check value then retain that value else take the check value, I=if $(S*W_i > \bar{x}_c, S * Wi, \bar{x}_c)$
- Final score: Sum the index value over all characters for genotypes. $F = \sum_{j=1}^{n} I_j$
- Selection: Final score with highest positive value will be selected as a best line.

R code

```
data=read.table(file.choose(), header=TRUE, row.names=1) # data file name stability test- folder R stability
install.packages("phenability")
library(phenability)
thsu(data, interaction=TRUE)
nahu(data, interaction=TRUE)
```

3. RESULTS AND DISCUSSION

Results of various non-parametric measures worked for several traits are presented as below.

i) Days to flowering

>thsu(data, interaction=TRUE)						
$ThSu	Hybrid	Mean	N1	N2	N3	N4
1	OKMSH-3	36.89	0.67	0.08	0.11	0.04
2	OKMSH-1	39.00	1.33	0.27	0.34	0.27
3	OKMSH-2	37.77	1.33	0.21	0.31	0.00
4	OKMSH-4	38.42	2.00	0.31	0.51	0.39
5	OKMSH-7	39.09	0.33	0.08	0.12	0.08
6	OKMSH-9	42.50	2.00	1.00	1.07	0.57
7	Shakthi(CC*)	45.00	2.33	1.17	1.85	1.40
8	AC-1685	41.67	2.33	0.47	0.74	0.08

*Commercial Check

>nahu(data, interaction=TRUE)

	Hybrid	Mean	S1	S2	S3	S6
1	OKMSH-3	36.89	0.33	1.00	1.00	1.33
2	OKMSH-1	39.00	1.33	4.33	0.50	0.50
3	OKMSH-2	37.77	0.00	5.33	2.88	1.53
4	OKMSH-4	38.42	2.00	10.33	3.96	1.65
5	OKMSH-7	39.09	0.33	0.33	0.40	0.40
6	OKMSH-9	42.50	1.33	9.33	0.10	0.20
7	Shakthi	45.00	2.33	14.33	0.09	0.18
8	AC-1685	41.67	0.33	14.33	3.34	1.31

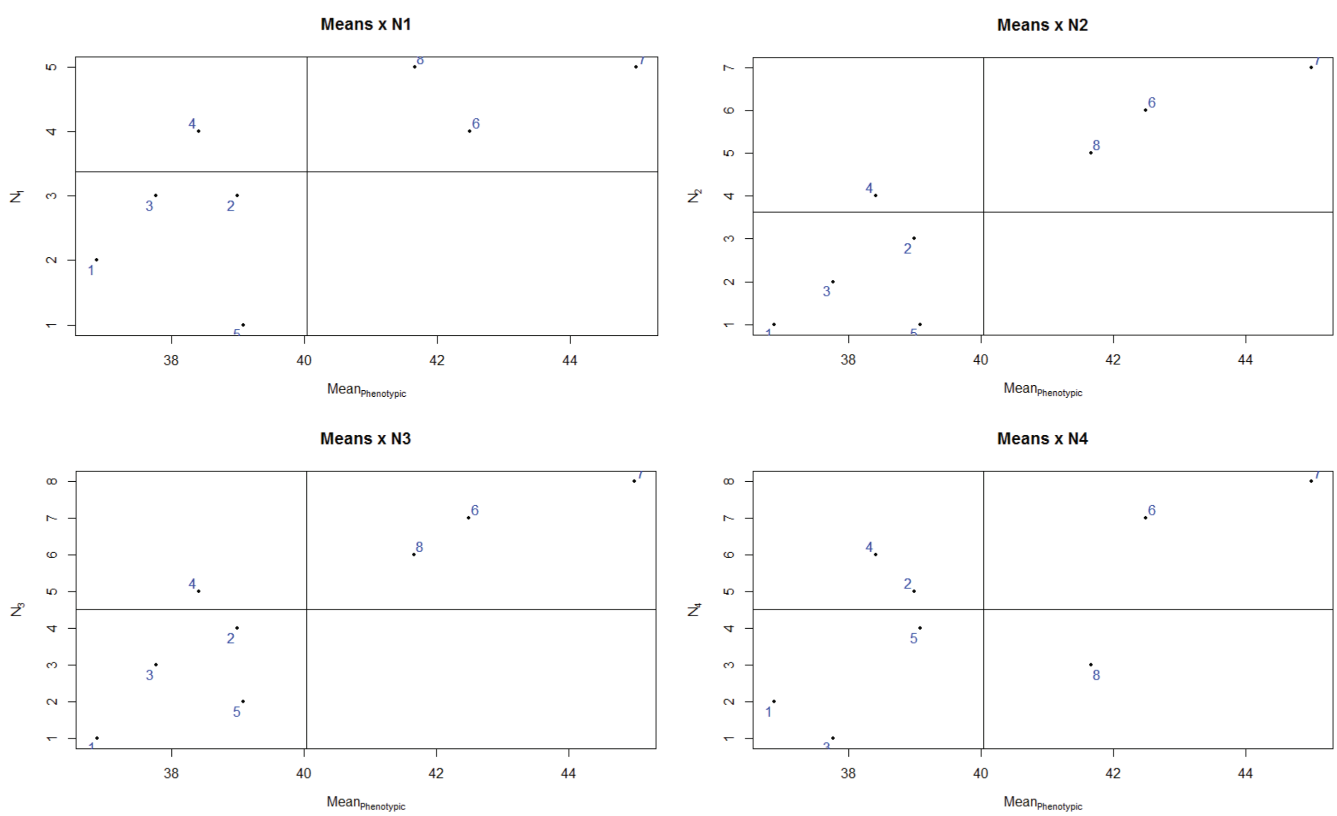


Fig. 2. Performance of N1-N4 measures based on mean phenotypic values for the character days to flowering

Table 1. Ranking of okra genotypes based on Thennarasu NP measure & Nasser and Huehn NP measure for the character days to flowering

	Hybrid	N1	N2	N3	N4	S1	S2	S3	S6
1	OKMSH-3	2	1	1	2	2	2	5	6
2	OKMSH-1	3	4	4	5	5	3	4	4
3	OKMSH-2	3	3	3	1	1	4	6	7
4	OKMSH-4	5	5	5	6	7	6	8	8
5	OKMSH-7	1	1	2	4	2	1	3	3
6	OKMSH-9	5	7	7	7	5	5	2	2
7	Shakthi	7	8	8	8	8	7	1	1
8	AC-1685	7	6	6	3	2	7	7	5

It may be observed that the hybrid OKMSH-3 has performed consistently well across most of the measures, followed by OKMSH-7 for days to flowering. Graphical representation of mean phenotypic value against the ranked measures (Fig.2) depicted also indicated pictorially the results presented in Table 1 for the measures N1 to N4. Similar pictorial representation for other measure/traits the ranks depicted in Table 1.

ii) Fruit weight

>thsu(data, interaction=TRUE)

	Hybrid	Mean	N1	N2	N3	N4
1	OKMSH-3	24.57	2.00	0.29	0.47	0.33
2	OKMSH-1	25.56	0.67	0.11	0.16	0.11
3	OKMSH-2	25.65	1.33	0.27	0.32	0.13
4	OKMSH-4	24.90	2.33	0.67	0.68	0.48
5	OKMSH-7	27.30	1.67	0.83	0.77	0.38
6	OKMSH-9	28.20	1.00	1.00	0.94	0.25
7	Shakthi	26.49	2.33	0.33	0.58	0.38
8	AC-1685	25.76	1.00	0.20	0.26	0.21

>nahu(data, interaction=TRUE)

	Hybrid	Mean	S1	S2	S3	S6
1	OKMSH-3	24.57	2.00	12.00	4.67	2.00
2	OKMSH-1	25.56	0.67	1.33	0.05	0.21
3	OKMSH-2	25.65	0.67	4.00	1.61	0.96
4	OKMSH-4	24.90	2.33	16.33	3.64	1.52
5	OKMSH-7	27.30	1.00	6.33	0.42	0.42
6	OKMSH-9	28.20	0.33	2.33	0.09	0.17
7	Shakthi	26.49	2.00	14.33	7.82	2.36
8	AC-1685	25.76	1.00	2.33	0.28	0.40

It may be observed that the hybrid OKMSH-1 has performed consistently well across most of the measures, followed by AC-1685 for fruit weight.

Table 2. Ranking of okra genotypes based on Thennarasu NP measure & Nasser and Huehn NP measure for the character days to fruit weight

	Hybrid	N1	N2	N3	N4	S1	S2	S3	S6
1	OKMSH-3	6	4	4	5	6	6	7	7
2	OKMSH-1	1	1	1	1	2	1	1	2
3	OKMSH-2	4	3	3	2	2	4	5	5
4	OKMSH-4	7	6	6	8	8	8	6	6
5	OKMSH-7	5	7	7	6	4	5	4	4
6	OKMSH-9	2	8	8	4	1	2	2	1
7	Shakthi	7	5	5	6	6	7	8	8
8	AC-1685	2	2	2	3	4	2	3	3

iii) Yield

>thsu(data, interaction=TRUE)

$ThSu						
	Hybrid	Mean	N1	N2	N3	N4
1	OKMSH-3	20.82	1.33	1.33	1.27	1.00
2	OKMSH-1	15.35	0.67	0.11	0.14	0.11
3	OKMSH-2	16.12	2.00	0.29	0.42	0.11
4	OKMSH-4	16.60	1.67	0.33	0.50	0.08
5	OKMSH-7	19.46	2.33	1.17	1.85	1.40
6	OKMSH-9	16.38	2.00	0.50	0.61	0.38
7	Shakthi	15.38	1.33	0.19	0.26	0.15
8	AC-1685	15.50	2.00	0.40	0.44	0.35

>nahu(data, interaction=TRUE)

	Hybrid	Mean	S1	S2	S3	S6
1	OKMSH-3	20.82	1.33	4.33	0.09	0.17
2	OKMSH-1	15.35	0.67	1.00	2.67	1.33
3	OKMSH-2	16.12	0.67	9.33	4.67	2.00
4	OKMSH-4	16.60	0.33	7.00	0.57	0.57
5	OKMSH-7	19.46	2.33	14.33	0.09	0.18
6	OKMSH-9	16.38	1.67	10.33	1.00	0.71
7	Shakthi	15.38	1.00	4.33	0.29	0.57
8	AC-1685	15.50	2.00	9.33	2.60	1.40

Table 3. Ranking of okra genotypes based on Thennarasu NP measure &Nasser and Huehn NP measure for the character yield

	Genotypes	N1	N2	N3	N4	S1	S2	S3	S6
1	OKMSH-3	2	8	7	7	5	2	1	1
2	OKMSH-1	1	1	1	2	2	1	7	6
3	OKMSH-2	5	3	3	2	2	5	8	8
4	OKMSH-4	4	4	5	1	1	4	4	3
5	OKMSH-7	8	7	8	8	8	8	2	2
6	OKMSH-9	5	6	6	6	6	7	5	5
7	Shakthi	2	2	2	4	4	2	3	3
8	AC-1685	5	5	4	5	7	5	6	7

It may be observed that the hybrid OKMSH-1 has performed consistently well across most of the measures for Yield.

iv) Incidence of YVMV

>thsu(data, interaction=TRUE)

$ThSu						
	Hybrid	Mean	N1	N2	N3	N4
1	OKMSH-3	8.94	0.33	0.06	0.08	0.00
2	OKMSH-1	11.11	1.33	0.24	0.38	0.27
3	OKMSH-2	5.44	2.00	0.31	0.39	0.30
4	OKMSH-4	11.20	2.00	0.36	0.61	0.43
5	OKMSH-7	8.26	1.00	0.18	0.29	0.21
6	OKMSH-9	12.86	2.00	0.67	0.87	0.11
7	Shakthi	9.20	0.33	0.07	0.10	0.00
8	AC-1685	81.59	2.33	2.33	3.30	0.00

>nahu(data, interaction=TRUE)

	Hybrid	Mean	S1	S2	S3	S6
1	OKMSH-3	8.94	0.00	0.33	0.17	0.33
2	OKMSH-1	11.11	1.33	5.33	1.63	1.00
3	OKMSH-2	5.44	2.00	10.33	1.36	1.14
4	OKMSH-4	11.20	2.00	12.00	2.58	1.23
5	OKMSH-7	8.26	1.00	3.00	3.16	1.36
6	OKMSH-9	12.86	0.33	10.33	0.33	0.33
7	Shakthi	9.20	0.00	0.33	0.28	0.40
8	AC-1685	81.59	0.00	16.33	0.00	0.00

Table 4. Ranking of okra genotypes based on Thennarasu NP measure & Nasser and Huehn NP measure for the character incidence of YVMV

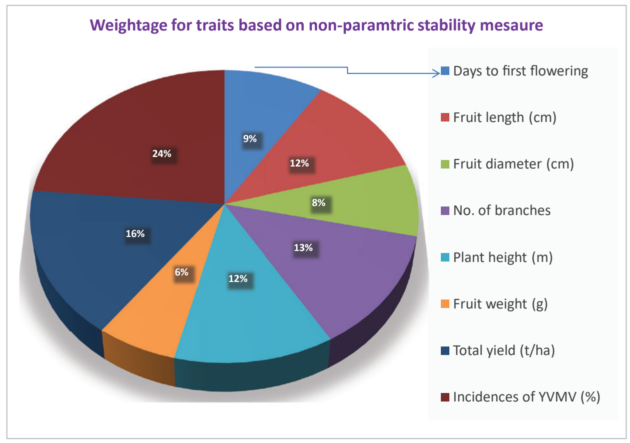
	Hybrid	N1	N2	N3	N4	S1	S2	S3	S6
1	OKMSH-3	1	1	1	1	1	1	2	2
2	OKMSH-1	4	4	4	6	6	4	6	5
3	OKMSH-2	5	5	5	7	7	5	5	6
4	OKMSH-4	5	6	6	8	7	7	7	7
5	OKMSH-7	3	3	3	5	5	3	8	8
6	OKMSH-9	5	7	7	4	4	5	4	2
7	Shakthi	1	2	2	1	1	1	3	4
8	AC-1685	8	8	8	1	1	8	1	1

It may be observed that the hybrid OKMSH-3 has performed consistently well across most of the measures, for the incidence of YVMV.

Similar analysis was carried out for the remaining 4 traits and it was noted that different hybrids were ranked best across different measures and there was no consistency. Accordingly, new index as discussed was adopted which was based on assigning derived weights (Fig.10) for all the traits and collective ranking based on all the traits.

Table 5. Results based on combined index for Okra (Non-Parametric approach for crop varietal release developed at ICAR-IIHR)

Name of the hybrid	IIHR NP (Venugopalan's NP measure)	
	Value	Rank
OKMSH-3	1034.48	**2**
OKMSH-1	239.52	6
OKMSH-2	1223.67	**1**
OKMSH-4	353.43	4
OKMSH-7	676.85	**3**
OKMSH-9	265.59	5
Shakthi	205.00	7
AC-1865	100.00	8



Weightage was based on data values recorded: more stable across replications /years for the evaluated genotypes, higher is the weightage.

Fig. 10. The weightage of various traits computed (for combined non-parametric index) Okra

Efficiency of combined index over the individual trait based index

Based on combined index, results revealed that the lines OKMSH 2, 3, 7 (in the same order) as superior with highest NP value as **1223.67(OKMSH 2) over all the evaluated traits.** This is probably due to the higher weight assigned to the trait incidence of YVMV, (incidence being least in OKMSH2) in addition to yield. Thus, there is a scope for releasing OKMSH2,3 and 7 as hybrids based on combined performance of all the characters.

4. CONCLUSION

In any crop improvement research, unpredictable environmental variation directly results in reduced gain due to selection, as the presence of G X E interaction would directly reduce the accuracy of prediction of genetic value. Stability solely based on single or 2-3 traits alone may not be sufficient, as the breeders expect that a hybrid /variety should also possess stability in desirable characters of other characters. A rank based non-parametric method has been suggested to identify a line/genotype evaluated over years as the best for varietal release simultaneously based on its superior performance over all traits, instead of one or two traits. Using the desired weights for individual traits arrived at based on its stability over years & within year replications, instead of assigning arbitrary weights, best lines were identified. Traits to be given reverse ranking (based on the improvement sought by the breeders over the existing cultivar) and direct ranking were also taken into consideration. It is suggested to make use of this method in varietal release / identification program and can be extended for MLT based varietal release.

ACKNOWLEDGEMENT

The authors wish to thank the Director, ICAR-IIHR, Bengaluru for providing all the facilities to carry out the research work and the anonymous referee for the critical comments which has led to a substantial improvement in the quality of the manuscript.

REFERENCES

Nassar, R. and Huehn, M. (1987) Studies on estimation of phenotypic stability; Tests of significance for non-parametric measures of phenotypic stability. *Biometrics,* **43,** 45-53.

Ravi, G.S.,Venugopalan, R., Padmini,K. and Gowda, D.M.(2013). Nonparametric measures for assessing yield stability in cucumber. *Int. J.of Agril. and Statist. Sci.***9** (1), 365-371.

Raiger, H. L. and Prabhakaran, V. T.(2001). A study on the performance of few non parametric stability measures.*Ind. J. Genetics*, **61**(1); 7-11.

Venugopalan,R. (2015). *Statistical Techniques in Quantitative Genetics for Fruit Crop Improvement Research.* Chapter 08, In *Fruit Breeding* (*Author*: M.R. Dinesh) under the series ‘Frontiers in Horticultural Science’, by K.V. Peter. ISBN NO. 9789383305513, New India Publishing Agency.

Venugopalan, R. and Veere Gowda, R. (2005). Stability analysis in Onion: A Statistical Look. *J. Indian Soc. Coastal Agric. Res.*, **23**(2), 123-29.

Venugopalan, R., and Pitchaimuthu M. (2009). Statistical Models for Stability Analysis In Watermelon, *J. Of Horti. Sci.*, **4**(2), 153-157.

Venugopalan, R. and Madhavi Reddy, K. (2010). Stability Analysis for Fruit Yield and Attributing Traits in Chilli.*Veg.Sci.***37** (2), 141-145.

Thennarasu, K.(1995). On certain non-parametric procedures for studying genotype-environment interactions and yield stability. *Indian J. Of Genet.,***60**, 433-439.

Annexure: R code for combined Index

```
rbd=read.csv(“D:\\1.Chaithra\\R\\np new.csv”)# take the standardized value
res1<-aov(A~as.factor(Rep)+as.factor(Trt),data=rbd)
res2<-aov(B~as.factor(Rep)+as.factor(Trt),data=rbd)
res3<-aov(C~as.factor(Rep)+as.factor(Trt),data=rbd)
res4<-aov(D~as.factor(Rep)+as.factor(Trt),data=rbd)
res5<-aov(E~as.factor(Rep)+as.factor(Trt),data=rbd)
res6<-aov(F~as.factor(Rep)+as.factor(Trt),data=rbd)
res7<-aov(G~as.factor(Rep)+as.factor(Trt),data=rbd)
res8<-aov(H~as.factor(Rep)+as.factor(Trt),data=rbd)
summary(res1)
summary(res2)
summary(res3)
summary(res4)
summary(res5)
summary(res6)
summary(res7)
summary(res8)
X<-c(0.479, 0.300, 0.619, 0.234, 0.259, 1.024, 0.146, 0.071)# mse of each variable
Y<-1/sqrt(X)
Z<-Y/sum(Y)*100
Z# weightage for each variable
CW<-as.matrix(read.table(file.choose(), header=TRUE, row.names=1))#CW=Check value and weight
MD<-as.matrix(read.table(file.choose(), header=TRUE, row.names=1))#MD=mean data
D<-cbind ((MD[,1]-CW[1,1]), (MD[,2]-CW[2,1]), (MD[,3]-CW[3,1]), (MD[,4]-CW[4,1]), (MD[,5]-CW[5,1]), (MD[,6]-CW[6,1]), (CW[7,1]-MD[,7]), (CW[8,1]-MD[,8]))
S<-cbind ((D[,1]/CW[1,1])*100, (D[,2]/CW[2,1]*100), (D[,3]/CW[3,1])*100, (D[,4]/CW[4,1]*100), (D[,5]/CW[5,1]*100), (D[,6]/CW[6,1]*100), (D[,7]/CW[7,1]*100), (D[,8]/CW[8,1]*100))
W1<-CW[1,2]
W2<-CW[2,2]
W3<-CW[3,2]
W4<-CW[4,2]
W5<-CW[5,2]
W6<-CW[6,2]
W7<-CW[7,2]
W8<-CW[8,2]
I1<-cbind((S[,1]*W1), (S[,2]*W2), (S[,3]*W3), (S[,4]*W4), (S[,5]*W5), (S[,6]*W6), (S[,7]*W7), (S[,8]*W8))
IA<-rbind((if (I1[1,1]>W1) {(I1[1,1])} else {(W1)}), (if (I1[2,1]>W1) {(I1[2,1])} else {(W1)}), (if (I1[3,1]>W1) {(I1[3,1])} else {(W1)}),(if (I1[4,1]>W1) {(I1[4,1])} else {(W1)}), (if (I1[5,1]>W1) {(I1[5,1])} else {(W1)}), (if (I1[6,1]>W1) {(I1[6,1])} else {(W1)}), (if (I1[7,1]>W1) {(I1[7,1])} else {(W1)}), (if (I1[8,1]>W1) {(I1[8,1])} else {(W1)}))
IB<-rbind((if (I1[1,2]>W2) {(I1[1,2])} else {(W2)}), (if (I1[2,2]>W1) {(I1[2,2])} else {(W2)}), (if (I1[3,2]>W1) {(I1[3,2])} else {(W2)}),(if (I1[4,2]>W1) {(I1[4,2])} else {(W2)}), (if (I1[5,2]>W1) {(I1[5,2])} else {(W2)}), (if (I1[6,2]>W1) {(I1[6,2])} else {(W2)}), (if (I1[7,2]>W1) {(I1[7,2])} else {(W2)}), (if (I1[8,2]>W1) {(I1[8,2])} else {(W2)}))
IC<-rbind((if (I1[1,3]>W3) {(I1[1,3])} else {(W3)}), (if (I1[2,3]>W1) {(I1[2,3])} else {(W3)}), (if (I1[3,3]>W1) {(I1[3,3])} else {(W3)}),(if (I1[4,3]>W1) {(I1[4,3])} else {(W3)}), (if (I1[5,3]>W1) {(I1[5,3])} else {(W3)}), (if (I1[6,3]>W1) {(I1[6,3])} else {(W3)}), (if (I1[7,3]>W1) {(I1[7,3])} else {(W3)}), (if (I1[8,3]>W1) {(I1[8,3])} else {(W3)}))
ID<-rbind((if (I1[1,4]>W4) {(I1[1,4])} else {(W4)}), (if (I1[2,4]>W1) {(I1[2,4])} else {(W4)}), (if (I1[3,4]>W1) {(I1[3,4])} else {(W4)}),(if (I1[4,4]>W1) {(I1[4,4])} else {(W4)}), (if (I1[5,4]>W1) {(I1[5,4])} else {(W4)}), (if (I1[6,4]>W1) {(I1[6,4])} else {(W4)}), (if (I1[7,4]>W1) {(I1[7,4])} else {(W4)}), (if (I1[8,4]>W1) {(I1[8,4])} else {(W4)}))
```

```
IE<-rbind((if (I1[1,5]>W5) {(I1[1,5])} else {(W5)}), (if (I1[2,5]>W1) {(I1[2,5])} else {(W5)}), (if (I1[3,5]>W1) {(I1[3,5])} else {(W5)}),(if (I1[4,5]>W1) {(I1[4,5])} else {(W5)}), (if (I1[5,5]>W1) {(I1[5,5])} else {(W5)}), (if (I1[6,5]>W1) {(I1[6,5])} else {(W5)}), (if (I1[7,5]>W1) {(I1[7,5])} else {(W5)}), (if (I1[8,5]>W1) {(I1[8,5])} else {(W5)}))
IF<-rbind((if (I1[1,6]>W6) {(I1[1,6])} else {(W6)}), (if (I1[2,6]>W1) {(I1[2,6])} else {(W6)}), (if (I1[3,6]>W1) {(I1[3,6])} else {(W6)}),(if (I1[4,6]>W1) {(I1[4,6])} else {(W6)}), (if (I1[5,6]>W1) {(I1[5,6])} else {(W6)}), (if (I1[6,6]>W1) {(I1[6,6])} else {(W6)}), (if (I1[7,6]>W1) {(I1[7,6])} else {(W6)}), (if (I1[8,6]>W1) {(I1[8,6])} else {(W6)}))
IG<-rbind((if (I1[1,7]>W7) {(I1[1,7])} else {(W7)}), (if (I1[2,7]>W1) {(I1[2,7])} else {(W7)}), (if (I1[3,7]>W1) {(I1[3,7])} else {(W7)}),(if (I1[4,7]>W1) {(I1[4,7])} else {(W7)}), (if (I1[5,7]>W1) {(I1[5,7])} else {(W7)}), (if (I1[6,7]>W1) {(I1[6,7])} else {(W7)}), (if (I1[7,7]>W1) {(I1[7,7])} else {(W7)}), (if (I1[8,7]>W1) {(I1[8,7])} else {(W7)}))
IH<-rbind((if (I1[1,8]>W8) {(I1[1,8])} else {(W8)}), (if (I1[2,8]>W1) {(I1[2,8])} else {(W8)}), (if (I1[3,8]>W1) {(I1[3,8])} else {(W8)}),(if (I1[4,8]>W1) {(I1[4,8])} else {(W8)}), (if (I1[5,8]>W1) {(I1[5,8])} else {(W8)}), (if (I1[6,8]>W1) {(I1[6,8])} else {(W8)}), (if (I1[7,8]>W1) {(I1[7,8])} else {(W8)}), (if (I1[8,8]>W1) {(I1[8,8])} else {(W8)}))
IV<-cbind(IA, IB, IC, ID, IE, IF, IG, IH) #IV= Index value
FS<-c(rowSums(IV)) # FS= final score
rank(-FS)
```

Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 74(1) 2020 77–84

Ontology Learning Algorithm for Development of Ontologies from Taxonomic Text and USDA Soil Taxonomy Ontology

Chandan Kumar Deb[1], Sudeep Marwaha[1] and R.N. Pandey[2]
[1]*ICAR-Indian Agricultural Statistics Research Institute, New Delhi*
[2]*ICAR-Indian Agricultural Research Institute, New Delhi*

Received 20 May 2017; Revised 20 June 2019; Accepted 29 June 2019

SUMMARY

Web based software use ontologies to keep the web one step ahead from the conventional web. Ontology based software architecture makes the platform suitable to work together for human as well as machine. By means of Ontology the unstructured knowledge is easily converted into the structured one. Soil Taxonomy Ontology developed for USDA soil taxonomy by Das (2010) and Das *et al.* (2012) for soil orders available in India is available to only sub group level. In this work developed knowledgebase is used to develop web based software with N-tier architecture and the ontology has been extended up to family and series level. It also covers all the twelve order of USDA Soil Taxonomy and provides the state wise series description of the soil. Search module of the software provides the exclusive search of the soil taxonomy and the edit module provide the facility to add, delete and edit facility of the ontology information. Additionally we have developed an algorithm for automated ontology learning from the taxonomic texts with a case study of soil taxonomy.

Keywords: Semantic web, Ontology, N-tier architecture, Ontology learning.

1. INTRODUCTION

Ontology is the heart of the semantic web and also acts in synergy with software agent and semantic web [Berners-Lee *et al.,* 2001]. Ontology helps in many ways to better describe the information of the web and their internal relationships. The main building block of the ontology is the statements. Statement defines concepts, relationships and imposes constraints to the concepts. Conceptually, this is very similar to the database schema or an object oriented class diagram. Across applications, communication can easily be achieved with the help of inbuilt ontology in the application. Although the scratch building of ontology is a very difficult task but once it is built; it can easily be extended and reused extensively.

The classic examples of developed ontologies are Gene Ontology [Gene Ontology Consortium (2000)] and Plant Ontology [Plant Ontology Consortium (2002)]. AmiGO functions as Browsing and searching tool for retrieving the data in Gene Ontology. Taxonomy has a great correspondence with the ontology. A methodology for conversion of taxonomies to ontology was proposed by Bedi and Marwaha (2004). The proposed methodology is tested and implemented for a pilot soil ontology using the IEEE standard Web Ontology Language (OWL) and protégé 2.1 OWL plug-in. OWL is the W3C recommendation for describing Ontology [Dean *et al.*, 2003]. Ontology-based intelligent retrieval system for soil knowledge [Ming *et al.*, 2009] is a system which searches documents related to soils by using soil domain ontology. This system retrieves information like "Relationship between Laterite soil and air pollution". Ontology Based Expert System [Bedi and Marwaha, 2005 and Marwaha, 2008] provides facilities to diagnose diseases and identify insects.

Soil Taxonomy Ontology has been built [Das (2010) and Das *et al.* (2012)] for USDA soil Taxonomy

Corresponding author: Chandan Kumar Deb
E-mail address: chandan.iasri@gmail.com

[USDA, NRCS (2010)] based on soil morphology that can be observed and measured in the field. In his work, a detailed study of the USDA soil taxonomy can be done by a given query interface, but his work didn't cover the taxonomy in totality. It only covered the seven orders among the twelve order of the USDA Soil Taxonomy. It also didn't cover the hierarchy up to the family and series level of the soil taxonomy. The developed Soil Taxonomy Ontology contains the information up to the sub group levels i.e. the family and series of the taxonomic hierarchy of the developed seven orders of soil ontology.

Under the present research work the Soil Taxonomy Ontology has been extended in two ways. Firstly, the existing seven orders are extended to the series level and secondly, the remaining five orders of the soil are also added to the taxonomy up to the subgroup level. All the user privileges remains intact in this research work and additionally two module namely state wise series description module and information edit module has been incorporated to the system. To overcome the shortcomings of the manual ontology building we have developed an algorithm for automated Ontology Learning with a case study of Soil Ontology.

2. MATERIAL AND METHODS

The Soil Taxonomy Ontology is a web based software with N-tier architecture. Entire application is developed based on two main tasks. First task is development of the user interface or the front end and the second task is the development of the back end which consists of database and knowledge base. These two ends of the software are bridged up by different java API (application programming interface).

2.1 Architecture of the Software

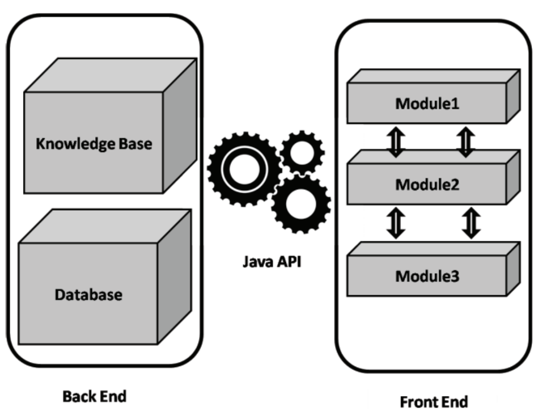


Fig. 1. Architecture of the Software

2.2 Schematic Process Flow of Development of the Software

Total process of the software development is described by the schematic diagram given in the Figure 2.2. Here we have used the USDA soil taxonomy as an information source for the development of the ontology. This is fed into the process flow of the development. In this process flow the identification of ontological building block i.e. class, instance, properties etc has been identified. After that the population of the ontology is done and the user interface is developed.

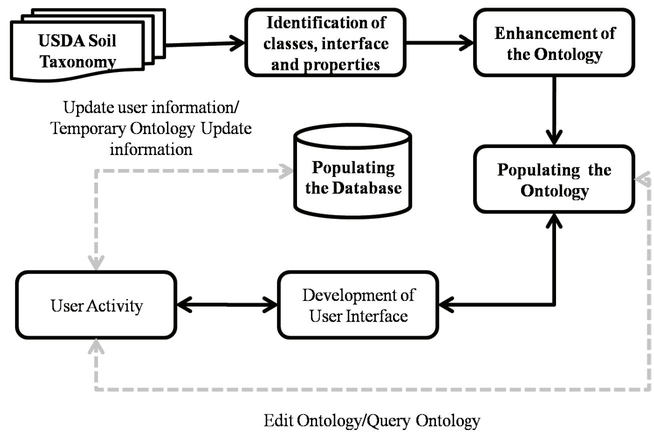


Fig. 2. Process flow of the development of soil taxonomy ontology

2.3 Identification of ontology Class, Individuals and Properties

2.3.1 Class Identification

Identification of class is the most important task of any ontology building. The names of the Orders, the formative element in the Order name, used as an identifier at lower categorical levels, derivation or source of the formative element and the mnemonicon for each Order. Each Suborder name consists of two syllables. The first is suggestive of the class (*i.e.* Suborder), and the second name of the Order is as reflected by the formative element (*e.g.* oll for Mollisol). Likewise, the names of Great Group are coined by prefixing one additional prefix (formative element) to the appropriate Suborder name. Subgroup names consist of the name of the appropriate Great group modified by one more adjective. Families, in this category, the intent has been to group the soils within a subgroup having similar physical and chemical properties that affect their responses to management and manipulation for use. For Series, the local name is used for the class name.

Examples:

Suppose the series ***Zaheerabad*** has the classification ***Clayey skeletal kaolinitic isohyperthermic Kandic Paleustalfs***

alfs is used for order Alfisols, **ustalfs** is used for the suborder ustalfs, **Paleustalfs** is used for the Great group, **Kandic Paleustalfs** is used for the sub group. So the family ***Clayey skeletal kaolinitic isohyperthermic Kandic Paleustalfs*** is the sub class of **Kandic Paleustalfs** class and the series ***Zaheerabad*** is the sub class of this family. In this manner a particular family or series are added to this ontology [Fig. 3.2].

In Soil Ontology the hierarchy is:

Order (Alfisols) Sub order (Ustalfs) Great group (Paleustalfs) Sub group (*e.g. Kandic Paleust*alfs) Family (*e.g. Clayey skeletal kaolinitic isohyperthermic Kandic Paleustalfs*) Series (*e.g. Zaheerabad*)

2.3.2 Individual Identification

According to the object oriented programming concepts, Individuals are the physical existence of the class. Like class identification for the ontology development, individual identification of every class is very important. For taxonomic class of the soil ontology we have used the same name as the class name. In case of property class like Basic_Property_Alfisols, we have created the property as individuals for order Alfisols.

In the same manner each and every classes are populated with different individuals.

2.3.3 Property Identification

Constructs and populates of Ontology are very much dependent on the property identification. Property is a very important component, because the two related class can only be joined by the property. On the basis of the related class, the property has been identified.

2.4 Tools and Technologies for Soil Ontology

The software development process is dependent in many ways on technology which is used for the development of the software. This software is developed using Java technology and total development process has been done on the Integrated Development Environment (IDE) Netbeans 6.9. All the development has been done through JDK 1.7 and additionally some API has been used to deal with the ontology.

In JEE 2.0 the web interface has been developed. The front end was developed by HTML, CSS and JavaScript. Apart from the core java class some of the programming is done through the JSP pages.

In the back end of the software has been divided into two parts. First part is the Database and Second part is the Knowledgebase. Behind the database development we used Microsoft SQL Server 2008

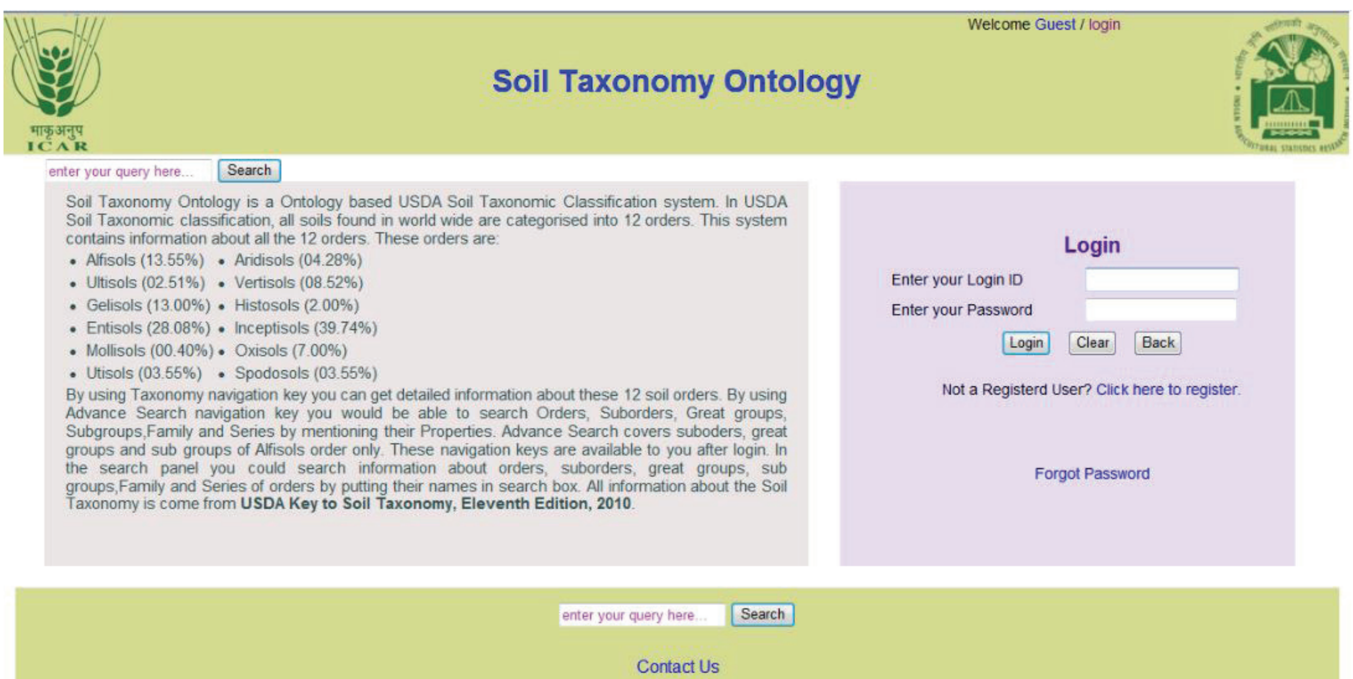


Fig. 3. Home Page of the Soil Taxonomy Ontology Software

and for the knowledgebase we used Protégé 3.4.6. Interaction of the front end with the back end has been done through the different java API. To connect to the database we have used the conventional JDBC Bridge but the connection as well as the interaction with the knowledgebase has been done through main java API namely; JENA, OWLSyntax, and ProtegeOWL. The second bridging process is literally known as the Semantic Web Framework Layer.

3. RESULTS AND DISCUSSION

3.1 Results of populated enhanced ontology

After the identification of class, interface and the properties of the domain i.e. USDA soil taxonomy, it is time to populate the Ontology with the real information. Some of the results are described below.

In this research work, we have worked on manual ontology building and also suggested an approach to make the process of ontology building automated. Although this work is done for 5 Orders, 20 Suborders, approx. 138 Great Group and approx. 793 Sub Groups.

Population of the ontology and proper tuning of soil taxonomy is one of the main objectives of this research work. The knowledge base of the Soil Taxonomy Ontology has been enhanced in many aspects. First, we have extended the existing seven orders up to the series level. Second, the population of the ontology up to the series level is done. Some of results of the populated ontology have been listed below.

i) The class has been extended for the world wide soil taxonomy.

ii) Property classes and their subclasses in soil ontology

iii) Property classes of family and series

iv) Classes for holding the geographical description of Series

v) Restriction applied to *has Basic Property* property

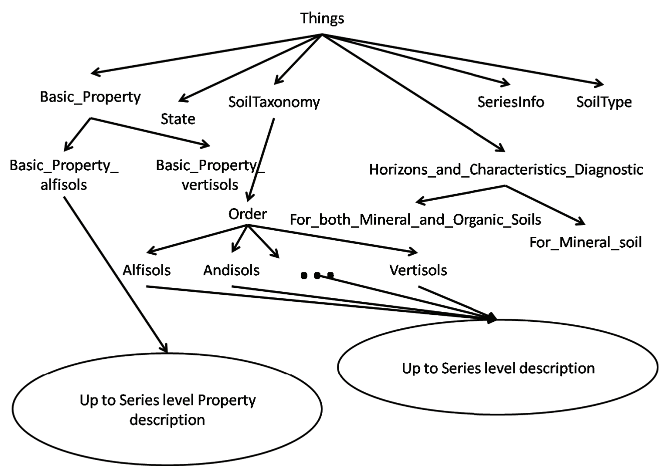


Fig. 3.1. Soil Ontology extended up to soil series

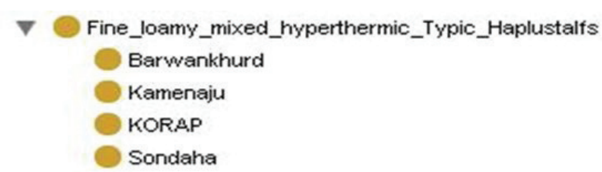


Fig. 3.2. Soil Ontology extended up to soil series in protégé

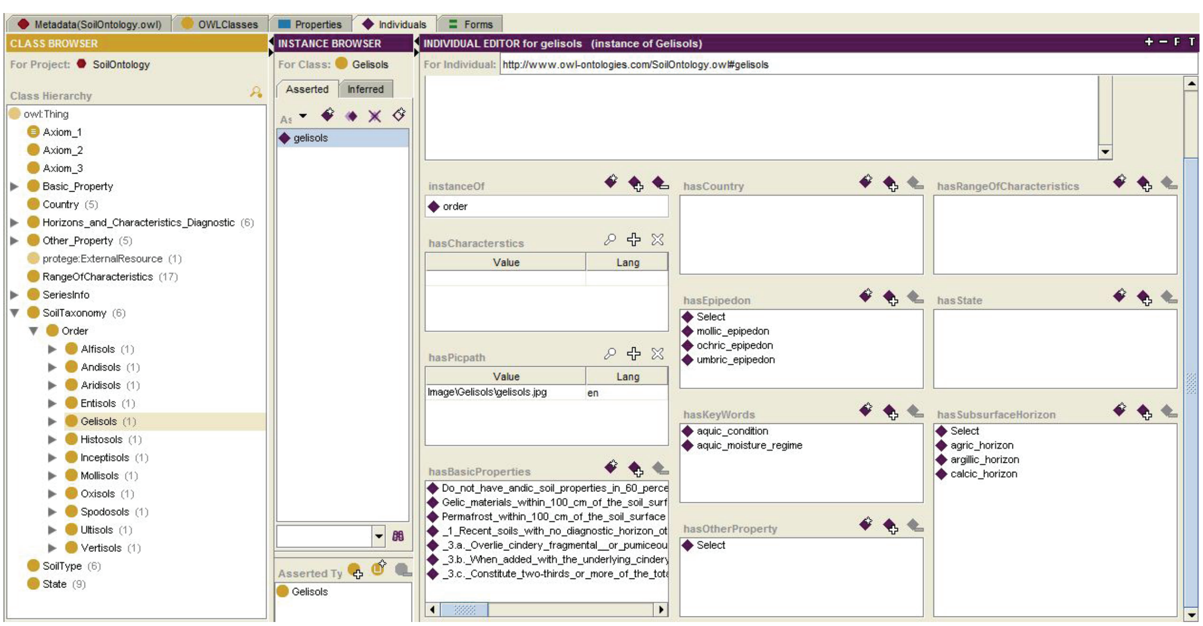


Fig. 3.3. Gelisols class with its individual gelisols with its properties and their corresponding values in Protégé OWL Plug-in

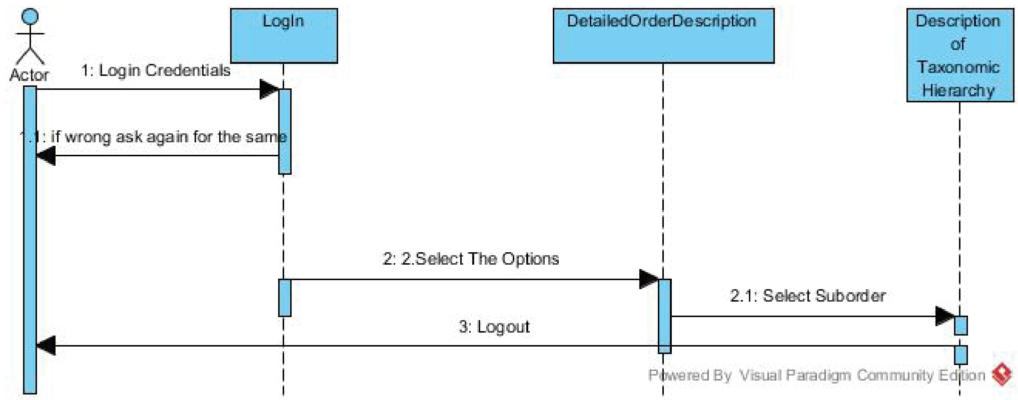


Fig. 3.1. Sequence diagram of Study of USDA Soil Taxonomy

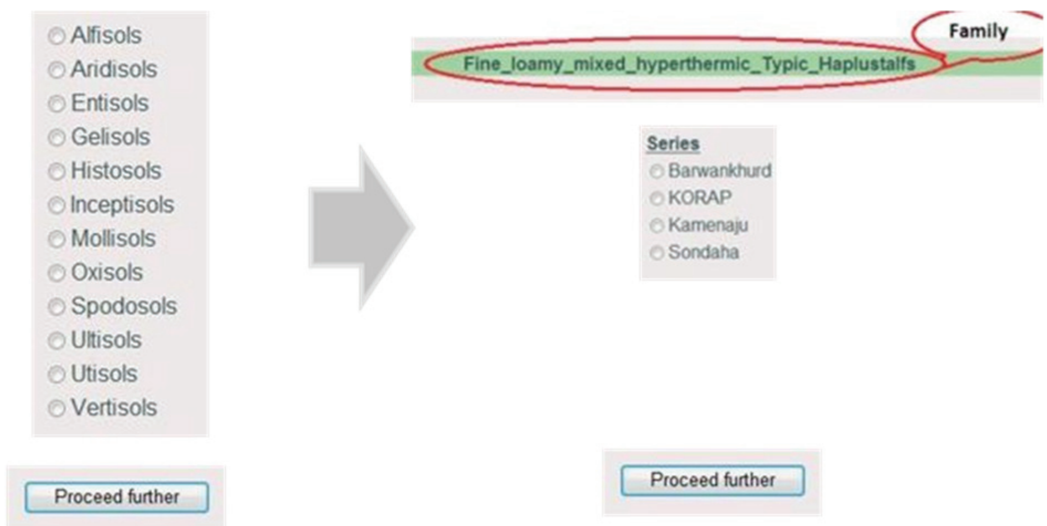


Fig. 3.4 Output of the study of Soil Taxonomy

3.2 Software facilitated the study of USDA soil taxonomy

The software provides the facility to a detailed study of the USDA soil taxonomy. It starts from the order and in a step by step manner it gives a detailed information of the selected category. In every step, it gives a details of the selected class and enlists the associated subclass. Fig 3.4 is the sequence diagram describing the steps of a detailed study of the soil taxonomy.

3.3 State wise series description of Indian soil series

For proper agricultural practice, the local information of the soil is very important. In the USDA Soil Taxonomy the series is the lowest hierarchy which is strongly coupled with the local soil description. One of the principle focus of this research work is to provide the series description of the soil. The software provides series information in two ways- firstly the taxonomic description which is available in "*Taxonomy*" tab of the software and secondly the state wise series description in "*Series*" tab of the software.

3.4 Classify newly found soil into proper hierarchy

Another powerful module of Soil Taxonomy Ontology software is totally deedicated to the searching of an existing hirarchy of the soil taxonomy. Figure 3.5 depicted the activity behind the search of the system. The search can be done through the simple search module or the advanced search module. The simple search is done by using the key words and the taxonomic term. The advanced search module is a relatively sophisticated one than the simple search module. The advanced search is done on the basis of specific information for any hierarchy (Order, suborder etc.) of the soil taxonomy. Both the search result produce a proper hierarchical format for proper understanding of the taxonomy.

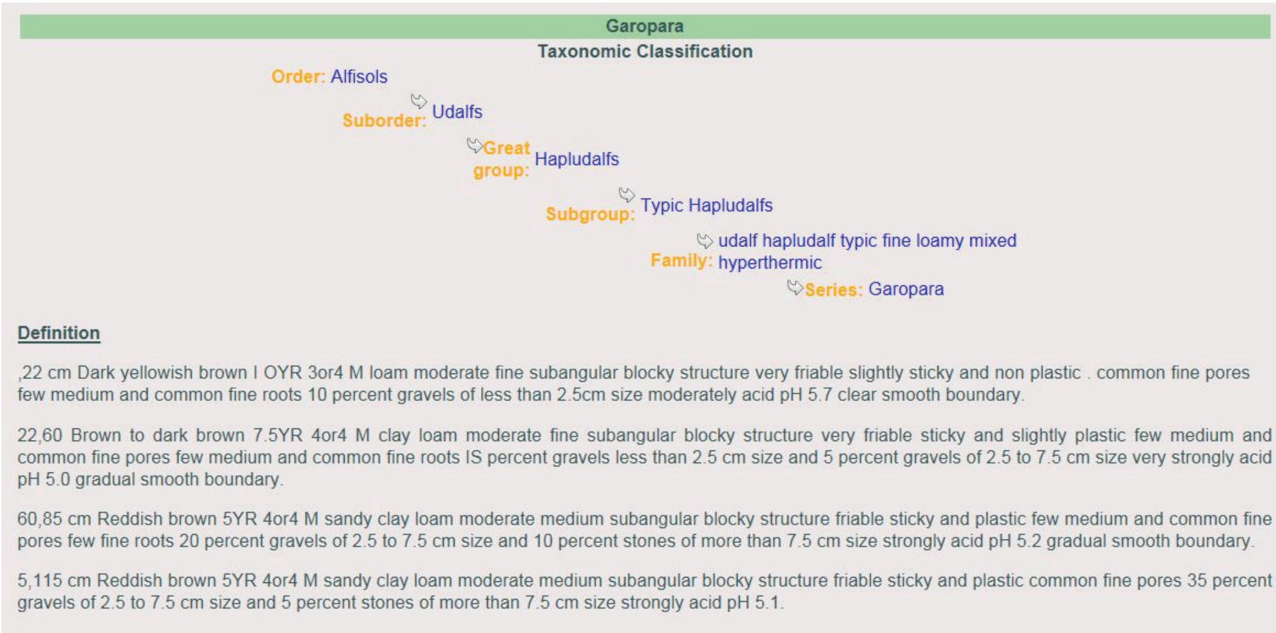


Fig. 3.6: Details study of Garopara Series by search

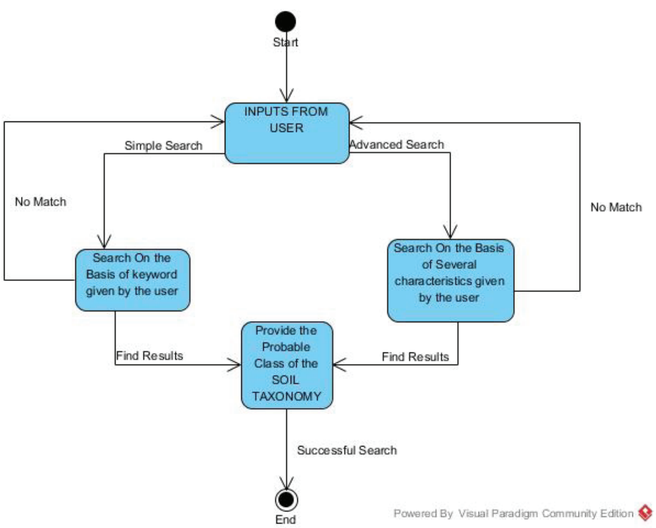


Fig. 3.5: Activity Diagram of Search Module

3.5 Software administrative functionality and ontology edit module

Web based software must have some administrative facility to combat many problems which appears after a long term use of the software. Cleaning of the garbage data, tuning of the software's necessary data also comes under the administrative functionality. This software has sign up facility to the user. Soil Taxonomy Ontology has three types of users. The administrator, domain experts and the general user are the users kind.

Among the three types, general user can only retrieve the information, use the simple and advanced search module and have the facility to classify newly found soil.

Like General user all the privileges are also available for the advanced user i.e The Domain Experts and the Administrator. Additionally they are involved in the Edit Ontology Module. The Domain Experts make the changes in the ontology. The change can be done in two ways first the new information which is already present in the ontology or edit the information which is already present in the ontology.

The Administrator is like super user of the system. Administrator can approve or disapprove the changes made by the domain experts.

4. AUTOMATED ONTOLOGY LEARNING ALGORITHM – CASE STUDY SOIL ONTOLOGY

As we have previously mentioned ontology building from the scratch is not only a difficult job but also it is very time consuming. This is very obvious that the taxonomic text is more structured than the plain text. We used this criterion to make the ontology development automated. We propose a methodology for ontology building through Natural Language Processing (NLP). We used the frame work of ontology learning proposed by Deb *et*. al. 2015.

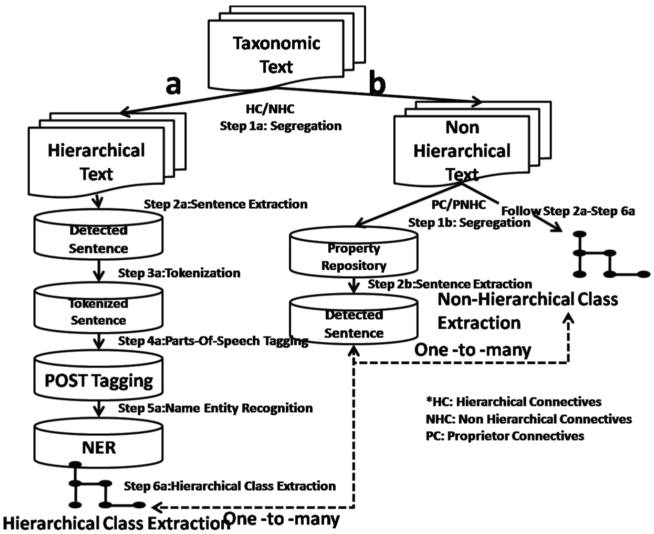


Fig. 4.1 Describes how the process of ontology learning may proceed

Step 1a and Step 1b: Segregation: This is the first step of ontology learning process. In this step total text or the part of the text which is under processing is divided into two parts i.e. the text contains the taxonomy and the text does not contains the taxonomy. The text is segregated on the basis of connectives present in the sentence. Connectives is a set that contains the key words that established connection between the objects available in taxonomic text.

e.g. “Udalfs are the Alfisols.” The sentence contains two object for ontology i.e “Udalfs” and “Alfisols”. “are the” established the linguistic connection between them. So for this text “are the” is one of the connectives.

Step 2a and Step 2b: Sentence Detection: Sentence detection is the next step of the ontology learning algorithm. Segment the total text into sentences for further task of NLP.

Step 3a: Tokenization: The sentence is further subdivided into words and single symbol called tokenization.

Step 4a: Parts-Of-Speech Tagging: In this task of NLP we find the proper noun for the identification of the taxonomic class of the taxonomic text.

Step 5a: Name Entity Recognition: After Step 4a Name Entity Recognition is very important. For detection of name the corpus can be built on the basis of the corresponding domain.

Step 6a: Hierarchical Class Recognition: In this step we have extracted the is-a relationship or the parent child relationship of the extracted name

First of all we provide the taxonomic text to the natural language processor. On the basis of connectives it will segregate the sentence into two parts hierarchical and non hierarchical text. Here the term connectives means a special set of words which describes the parent-child relationship in the text. In case of taxonomic text it is more prominent and easily usable. The above described separation of the sentence group makes the task of taxonomic relation extraction very easy.

After separation of the hierarchical and non hierarchical text the non hierarchical text further segregated on the basis another connectives i.e. proprietor connectives. This connective is used for the separation of properties for a particular class which is extracted from the hierarchical text or non hierarchical text.

On one hand we have the extracted taxonomic class and subclass and on the other hand we have extracted the properties. Both the extracted things are attached through some semi automated process (Manual relation extraction and association rule extraction) and the Ontological structure will easily be built.

5. CONCLUSION

In this work we have developed web based software with N-tier architecture with Soil Ontology as a knowledgebase. The software provides the facility for a detailed study of the USDA Soil Taxonomy up to the series level. The developed software has the state wise series description facility which can be used for the local soil description that can easily be used in the agricultural practice. It also provides the editing facility of the existing ontology. We have also developed an algorithm for automated ontology learning.

ACKNOWLDGEMENTS

The first author gratefully acknowledges the INSPIRE Fellowship provided by Department of Science and Technology, New Delhi.

REFERENCES

Berners-Lee T, Hendler J and Lassila O. (2001). The Semantic Web. *Scientific American,* 284 (5): 34-43.

Bedi, P., Marwaha, S. (2004).Designing Ontologies from Traditional Taxonomies, *In the Proceedings of International Conference on Cognitive Science*, Allahabad, India.

Bedi, P. and Marwaha, S. (2005). Framework for Ontology Based Expert Systems: Disease and Pests Identification in Crops-A case study, in Proceedings of the International Conference of Artificial Intelligence. (ICAI-05): 256-259.

Clark, K. (2008). *SPARQL Protocol for RDF*, *W3C Recommendation*, http://www.w3.org/TR/rdf-sparql-protocol/.

Das, M. (2010). Building and querying soil ontology for agriculture. Unpublished M.Sc. Thesis., IARI, New Delhi.

Das, M., Malhotra, P.K., Marwaha, Sudeep and Pandey, R.N. (2012). Building and Querying Soil Ontology. *J. Indian. Soc. Ag. Stat.* 66(3): 459-464.

Dean, M., Connolly, D., Harmelen, F.V., Hendler, J., Horrocks, I., McGuinness, D.L., Patel-Schneider, P.F. and Stein, L.A. (2003). *OWL Web Ontology Language Reference, W3C Working Draft.* http://www.w3.org/TR/2003/WD-owl-ref-20030331.

Deb, C.K., Marwaha, S., Arora, A., & Das, M. (2018). A framework for ontology learning from taxonomic data. In Big Data Analytics (pp. 29-37). Springer, Singapore.

Gene Ontology Consortium. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25: 25-29.

Maji, A.K., Baruah U., Dubey P.N., Butte P.S., Verma T.P, Shilu K. and Angami V. (2004a). Soil Series of Nagaland, NBSS Publ. No.109. NBSS&LUP, Nagpur,pp126.

Maji, A.K., Sarkar D., Baruah U., Bhaskar B.P., Sen T.K, Vadivelu S. and Gajbhiye K.S. (2004b). Soil Series of Assam, NBSS Publ. No.101. NBSS&LUP, Nagpur, pp229.

Maji, A.K., Sarkar D., Baruah U., Sen T.K and Patil V.P. (2006). Soil Series of Manipur, NBSS Publ. No.134. NBSS&LUP, Nagpur, pp53.

Marwaha, S. (2008). Temporal Extension to Ontologies for Semantic Web enabled Systems. Unpublished Ph. D. Thesis, University of Delhi.

Ming, Z., Qingling, Z., Dong, T., Ping, Q., and Xiaoshuan, Z. (2009). Ontology-based intelligent retrieval system for soil knowledge. *WSEAS Trans. Info. Sci. and App. 6, 7*. http://portal.acm.org/citation.cfm?id=1639432.

NBSS & LUP Staff (2006). NBSS Publ. No.136. Soil Series of Kerala, NBSS &LUP, Nagpur, pp136.

Plant Ontology Consortium. (2002). Plant Ontology Consortium. http://www.plantontology.org/.

Sahoo, A.K, Sarkar, D. and Gajbhiye, K.S. (2004). Soil Series of Bihar, Publ. No.98 NBSS &LUP, Nagpur, pp289.

Sarkar, D., Gajbhiye, K.S, Sahoo A.K. and Sah, K.D. (2005). Soil Series of Orissa, NBSS Publ. No.119 NBSS&LUP, Nagpur, pp254.

Sarkar, D., Dubey P.N., Ray S.K., Baruah U., Sehgal, J. and Bhattacharyya, T. (2004). Soil Series of Tripura, NBSS Publ. No.111 NBSS &LUP, Nagpur, pp115.

Singh, R.S., Baruah U., Sarkar, D, Butte P.S., Gajbhiye, K.S. (2005). Soil Series of Meghalaya, NBSS Publ. No.121 NBSS&LUP, Nagpur, pp86.

Singh, R.S., Singh, S.K., Jain, B.L., Shyampura, R.L. and Gajbhiye, K.S. (2002). Soil Series of Rajasthan, NBSS Publ. No.95 NBSS&LUP, Nagpur, pp364.

Smith, M., Welty, Chris and McGuinness, Deborah L. (2004). *OWL Web Ontology* Language Guide (W3C Recommendation). http://www.w3.org/TR/owl-guide-20040210/.

USDA, NRCS. (2010). *Keys to Soil Taxonomy*, Eleventh Edition.

भारतीय कृषि सांख्यिकी संस्था की पत्रिका
अंक 74 खंड 1 जनवरी 2020 85-89

हिन्दी परिशिष्ट: इस खंड में प्रकाशित शोधपत्रों के सारांश

अंक 74 जनवरी 2020 खंड 1

कृषि सांख्यिकी: सिद्धांत एवं अनुप्रयोग अनुक्रमणिका

संगणक अनुप्रयोग

कई प्रतिक्रियाओं के लिए स्तरीकृत याद्दच्छिक प्रतिक्रिया प्रतिमान

रघुनाथ अर्नब एवं डी.के. शानगोडोयिन

बोट्सवाना विश्वविद्यालय, बोट्सवाना

संवेदनशील विशेषताओं पर डेटा एकत्र करने के लिए याद्दच्छिक प्रतिक्रिया (आरआर) तकनीकों का उपयोग किया जाता है। अब्देल्फतह व मजलूम (2015) ने समतामूलक अध्ययन के आधार पर स्तरीकृत नमूने और दावे के लिए दो डेक के आधार पर ओडूमड़े व सिंह (2009) की आरआर तकनीकों को विकसित किया है जो कि उनके प्रस्तावित अनुमानक से अधिकतर स्तिथियों में बेहतर प्रदर्शन करते हैं। इस लेख में हमने ओडूमड़े व सिंह (2009) व अब्देल्फतह व अन्य (2011) के प्रत्येक वैकल्पिक अनुमानकों का प्रस्ताव किया है और स्तरीकृत नमूने के लिए अब्देल्फतह व मजलूम (2015) की आरआर तकनीक का प्रयोग किया है। प्रस्तावित अनुमानक वर्तमान अनुमानकों की तुलना में अधिक कुशल पाए गए हैं। बढ़ी हुई क्षमताओं के अतिरिक्त प्रस्तावित अनुमानक अनुपात के अनुमानकों, भिन्नताओं और भिन्नताओं के निष्पक्ष अनुमानकों के लिए सरल भाव रखते हैं।

भारत में जलवायु लचीला, उच्च उपज और स्थिर गन्ना जीनोटाइप

राजेश कुमार, ए.डी. पाठक एवं बक्शी राम

भा.कृ.अ.प. - भारतीय गन्ना अनुसंधान संस्थान, लखनऊ

गन्ने पर अखिल भारतीय समन्वित अनुसंधान परियोजना (AICRP) के उन्नत रूपांतर ट्रेल्स के दीर्घकालिक डेटा विश्लेषण के आधार पर, नौ जीनोटाइप की पहचान की गई है जिसमें 2012 से 2018 तक पर्यावरणीय परिस्थितियों में प्रतिकूल बदलाव के प्रति उच्च उपज स्थिरता और कम संवेदनशीलता के गुण हैं। प्रारंभिक समूह से दो सह 10024 और सह 11001 और प्रायद्वीपीय क्षेत्र के मध्य देर से समूह से दो सीओएम 11086 और सह 08009। ईस्ट कोस्ट जोन में, केवल एक प्रारंभिक परिपक्व जीनोटाइप, सीओए 13322 की पहचान की गई थी। उत्तर मध्य क्षेत्र से दो मध्य-देर के जीनोटाइप, सीओएच 08262 और सीओएच 09264 की पहचान की गई थी। इसी तरह उत्तर मध्य और उत्तर पूर्वी क्षेत्र से दो मध्य-देर के जीनोटाइप, CoSe 11454, CoP 12438 की पहचान की गई। 163 जीनोटाइप्स में से, केवल CoSe 11454 तीनों चरित्र, CCS (t/ha), गन्ने की उपज (t/ha) और सुक्रोज (%) के लिए अत्यधिक स्थिर था। अन्य आठ केवल CCS (t/ha) और गन्ने की उपज (t/ha) के लिए अत्यधिक स्थिर थे। पारगमन कार्यक्रम में माता-पिता के रूप में इनका उपयोग उच्च उपज स्थिरता और पर्यावरणीय परिस्थितियों में प्रतिकूल बदलावों के प्रति कम संवेदनशीलता और उच्च स्थिरता और उच्च उपज मानदंड के गुणों के रूप में किया जा सकता है।

कुछ रूपांतरित और समग्र श्रृंखला अनुपात-प्रकार अनुमानकों के दो सहायक चरों का उपयोग करना

सर्वेश कुमार दुबे[1], बी.वी.एस. सिसोदिया[1] एवं मनोज कुमार शर्मा[2]

[1]नरेंद्र देव कृषि एवं प्रौद्योगिकी विश्वविद्यालय, फैजाबाद

[2]गिरि विकास अध्ययन संस्थान, लखनऊ

वर्तमान लेख में अध्ययन के तहत चर से संबंधित दो सहायक चरों पर जानकारी का उपयोग करते हुए परिमित जनसंख्या सर्वेक्षण नमूने में आबादी के कुछ रूपांतरित और समग्र श्रृंखला अनुपात प्रकार के अनुमानकर्ताओं से निपटा है। उनके पूर्वाग्रह और एमएसई व्युत्पन्न हैं। वास्तविक डेटा का उपयोग करके साहित्य में कई मौजूदा अनुमानकों की तुलना में प्रस्तावित अनुमानकों की सापेक्ष दक्षता की जांच की गई है। यह पाया गया है कि प्रस्तावित अनुमानकों ने मौजूदा श्रृंखला अनुपात प्रकार के अनुमानकों और समग्र श्रृंखला अनुपात प्रकार के अनुमानकर्ताओं से बेहतर प्रदर्शन किया है।

दो चरण के नमूने के तहत परिमित जनसंख्या का उत्पाद प्रकार अंशांकन अनुमानक

अंकुर बिस्वास, कौस्तव आदित्य, यू.सी. सूद एवं प्रदीप बसाक

भा.कृ.अ.प. - भारतीय कृषि सांख्यिकी अनुसंधान संस्थान, नई दिल्ली

Deville और Särndal (1992) द्वारा प्रस्तावित कैलिब्रेशन दृष्टिकोण सर्वेक्षण नमूने में कुशलतापूर्वक सहायक जानकारी का उपयोग करने के लिए एक लोकप्रिय

तकनीक है। इस अध्ययन में, परिमित आबादी के अंशांकन आकलनकर्ताओं को दो चरण के नमूने के डिजाइन के तहत विकसित किया गया है जिसमें अनुमानक के विचरण और प्रसरण के संबंधित अनुमानक के साथ-साथ नमूनाकरण डिजाइन है। यह माना जाता है कि जनसंख्या स्तर की जटिल सहायक जानकारी चयन के दूसरे चरण में उपलब्ध है और अध्ययन चर उपलब्ध सहायक जानकारी से विपरीत है। प्रस्तावित अंशांकन आकलनकर्ताओं का मूल्यांकन एक सिमुलेशन अध्ययन के माध्यम से किया गया था और यह पाया गया था कि सभी प्रस्तावित उत्पाद प्रकार के अंशांकन आकलनकर्ता हॉर्विट्ज-थॉम्पसन अनुमानक की तुलना में बेहतर प्रदर्शन करते हैं और साथ ही दो चरण नमूना डिजाइन के तहत कुल जनसंख्या का उत्पाद अनुमानक भी है।

बहु-गियर मत्स्य पालन के लिए स्टॉक की स्थिति का आकलन करने के लिए शेफर प्रोडक्शन मॉडल का बेजियन राज्य-स्थान कार्यान्वय

एल्डो वर्गीज, टी.वी. साथियानंदन, जे. जयसंकर, सोमी कुरीयाकोज, के.जी. मिनी एवं एम. मुक्ता

भा.कृ.अ.प. - केंद्रीय समुद्री मात्स्यिकी अनुसंधान संस्थान, कोच्चि

समुद्री मछली स्टॉक की स्थिति को जानना समुद्री मत्स्य संसाधनों की स्थायी फसल के लिए प्रबंधन रणनीतियों को विकसित करने के लिए अत्यंत महत्वपूर्ण है। इस बात के लिए एक व्यापक रूप से स्वीकृत दृष्टिकोण मछली पकड़ने और मछली पकड़ने के प्रयासों पर आधारित टाइम सीरीज डेटा का उपयोग करते हुए स्थायी फसल के स्तर को प्राप्त करना है जो फिशर के मॉडल जैसे कि बायोमास गतिकी का वर्णन करता है। भारत में, समुद्री मत्स्य जटिल बहु-प्रजाति की प्रकृति है, जहाँ विभिन्न प्रजातियों में मछली पकड़ने के गियर और प्रत्येक गियर हारवेस्ट द्वारा कई प्रजातियों को पकड़ा जाता है, जिससे प्रत्येक मछली की प्रजातियों के अनुरूप मछली पकड़ने का प्रयास करना मुश्किल हो जाता है। चूंकि गियर की क्षमता भिन्न होती है, इसलिए संसाधन को पकड़ने के लिए किए गए प्रयास को विभिन्न मछली पकड़ने के गियर द्वारा खर्च किए गए प्रयासों के योग के रूप में नहीं माना जा सकता है। इसलिए, यह स्टॉक मूल्यांकन मॉडल में उपयोग करने के लिए प्रयास मानकीकरण के महत्व की मांग करता है। इस पत्र में मछली पकड़ने के प्रयासों के मानकीकरण और फिशर उत्पादन मॉडल (बीएसएम) के बेजियन राज्य-अंतरिक्ष कार्यान्वयन का उपयोग करके मछली स्टॉक की स्थिति का आकलन करने के लिए एक पद्धति का वर्णन किया गया है। एक मोंटे-कार्लो आधारित विधि जिसका नाम कैच-मैक्सिमम सस्टेनेबल यील्ड (CMSY) है, का उपयोग लैंडिंग से मछली पालन के संदर्भ बिंदुओं का अनुमान लगाने और प्रजातियों की लचीलापन का उपयोग करके बायोमास के लिए एक प्रॉक्सी के लिए किया गया है। यह प्रक्रिया 1997-2018 के दौरान भारत के तटीय राज्य आंध्र प्रदेश, भारत से एकत्र किए गए भारतीय मैकेरल (रैस्त्रेलिगर कानागुरता) के आंकड़ों के साथ चित्रित की गई है। आंध्र प्रदेश के लिए भारतीय मैकेरल की अधिकतम सतत उपज (MSY) का अनुमान लगाया गया है। CMSY और BSM दोनों तरीकों के बीच तुलना की गई है और पाया गया है कि अनुमान करीबी समझौतों में हैं।

सीआरडी और आरबीडी सेट-अप में इष्टतम कोवरिएट डिजाइन की नई श्रृंखला

हिरण्मय दास[1], अनुरुप मजूमदार[2], मनोज कुमार[3] एवं दिकेश्वर निषाद[4]

[1]भा.कृ.अ.प. - भारतीय मृदा विज्ञान संस्थान, भोपाल

[2]बिधान चंद्र कृषि विश्वविद्यालय, नाडिया

[3]भा.कृ.अ.प. - केंद्रीय कृषि अभियांत्रिकी संस्थान, भोपाल

[4]पंडित शिव कुमार शास्त्री कृषि महाविद्यालय एवं अनुसंधान केंद्र, राजनंदगांव

सीआरडी सेट-अप में कोवरिएट मॉडल के लिए इष्टतम डिजाइन का अध्ययन ट्रॉय (1982 ए, 1982 बी) द्वारा शुरू किया गया था। दास एट अल (2003) अध्ययन के बाद और आरबीडी सेट-अप के लिए बढ़ा दिया गया। हाल ही में दास एट अल (2015) ने 'ऑप्टिमल कोवरिएट डिजाइन' पर एक पुस्तक प्रकाशित की। वर्तमान अध्ययन में, सीआरडी सेट-अप में आरबीआईडी सेट-अप में दो वैश्विक श्रृंखला और कॉर्बेट डिजाइन की एक नई श्रृंखला विकसित की गई है। CRD या RBD डिजाइनों में नए OCDs को आदेश 2 और 4. के केवल दो हैडमर्ड मेट्रिसेस की आवश्यकता होती है। CRD सेट-अप में विकसित वैश्विक

इष्टतम कोवरिएट डिजाइन में v (= 0; mod 4 या = 2; mod 4; उपचार की संख्या), और हैं। RBD सेट-अप में वैश्विक इष्टतम कोवरिएट डिजाइन की विकसित पहली श्रृंखला में किसी भी प्रतिकृति या ब्लॉक की संख्या के लिए उपचार नंबर v (= 1; mod 4) है, b और Hv और Hb के अस्तित्व पर निर्भर नहीं है। आरबीडी सेट-अप में वैश्विक इष्टतम कोवरिएट डिजाइन की दूसरी श्रृंखला को केवल एचवी के अस्तित्व की आवश्यकता होती है। कागज इष्टतम कोवरिएट्स के उदाहरणों से समृद्ध है। वर्तमान लेख में सभी विकसित इष्टतम कोवरिएट डिजाइन मौजूदा साहित्य में उपलब्ध नहीं हैं।

आरआरटी का उपयोग करके भारत के उत्तराखंड के कुमायूं क्षेत्र में स्थित वरिष्ठ माध्यमिक विद्यालयों के छात्रों के बीच मादक पदार्थों के व्यसन का अनुमान

नीरज तिवारी एवं तनुज कुमार पांडेय

कुमाऊं विश्वविद्यालय, एस.एस.जे. परिसर, अल्मोड़ा

हमने भारत के उत्तराखंड के कुमायूं क्षेत्र में स्थित वरिष्ठ नागरिकों के स्कूलों के पुरुष और महिला छात्रों के बीच नशीली दवाओं की संख्या का अनुमान लगाने के लिए यादृच्छिक प्रतिक्रिया तकनीक (आरआरटी) लागू किया। वरिष्ठ माध्यमिक के छात्रों के बीच नशा मुक्ति का अनुमान लगाने के लिए वार्नर के आरआरटी को लागू किया गया था। क्षेत्र। अध्ययन में अपनाई गई प्रक्रिया उत्तरदाताओं को पर्याप्त गोपनीयता प्रदान करती है और सर्वेक्षण के समय को कम करती है। कुल छात्रों में से 26% की अनुमानित संख्या 95% आत्मविश्वास अंतराल [20%, 32%] के साथ ड्रग्स की आदि पाई गई। छात्रों के मादक पदार्थों की लत और अकादमिक प्रदर्शन के बीच एक मजबूत संबंध पाया गया था। यह अध्ययन राज्य और केंद्र सरकारों में योजनाकारों के लिए अत्यधिक उपयोगी है जो क्षेत्र के स्कूल जाने वाले बच्चों के बीच मादक पदार्थों के व्यसन के गुरुत्वाकर्षण का आकलन करते हैं और बढ़ते नियंत्रण के तरीके खोजते हैं।

केरल में धान पर दीर्घकालिक उर्वरक प्रयोगों में पोषक तत्वों के उत्थान की गतिशीलता

वी.ए. जेसमा[1], टी.के. अजीथा[1], एस. कृष्णन[1], पी.पी. मूसा[2] एवं पी सिंधुमोले[1]

[1]बागवानी महाविद्यालय, वेलानिकारा

[2]आरएआरएस, पट्टाम्बी

वर्तमान अध्ययन अनाज उपज के द्वितीयक आंकड़ों पर आधारित है, जो खरीफ और रबी मौसम के दौरान RARS पट्टांबी में आयोजित चावल पर दीर्घकालिक उर्वरक प्रयोग (LTFE) पर AICRP से प्राप्त किया गया है। अध्ययन का उद्देश्य नॉनलाइन रिग्रेशन का उपयोग करते हुए चावल की अनाज उपज पर एन, पी और के के ऊपर पौधों के पोषक तत्वों के प्रभाव का अध्ययन करना था। द्विघात मॉडल दोनों मौसमों में उपज और पौधों के पोषक तत्वों के बीच संबंधों को पकड़ने में सक्षम था ।

अनुपात का एक विकल्प और स्तरीकरण के लिए दोहराई गई नमूनाकरण में परिमित जनसंख्या के उत्पाद प्रकार के अनुमानक

हिलाल ए. लोन[1], राजेश टेलर[2] एवं मेद राम वर्मा[3]

[1]राजकीय डिग्री महाविद्यालय, सोपोर, जम्मू व कश्मीर, भारत

[2]विक्रम विश्वविद्यालय, उज्जैन

[3]भा.कृ.अ.प. - भारतीय पशु चिकित्सा अनुसंधान संस्थान, इज्जतनगर

इस पत्र में हमने इगे और त्रिपाठी (1987) के अनुमानों का एक विकल्प प्रस्तावित किया है। पूर्वाग्रह और औसत चुकता त्रुटियों के लिए अभिव्यक्तियाँ सन्निकटन की पहली डिग्री तक प्राप्त की गई हैं। प्रस्तावित अनुमानकों की तुलना जनसंख्या के सामान्य निष्पक्ष अनुमानक से की गई है जो स्तरीकरण और अनुपात और उत्पाद प्रकार के अनुमानों के लिए डबल नमूने में ईगे और त्रिपाठी (1987) द्वारा दिए गए हैं। प्रस्तावित अनुमानकों के गुणों का न्याय करने के लिए एक अनुभवजन्य अध्ययन किया गया है।

बागवानी फसल वैराइटी भिन्नता के लिए गैर-पैरामीट्रिक स्थिरता दृष्टिकोण

आर. वेणुगोपालन, एम. पीतचिमुथु एवं एम. चौथ्रा

भा.कृ.अ.प. - भारतीय बागवानी अनुसंधान संस्थान, बेंगलुरु

स्थिरता के साथ युग्मित कई लक्षणों के लिए एक पंक्ति के प्रदर्शन के आधार पर फसल के varietal रिलीज के लिए एक गैर-पैरामीट्रिक स्थिरता सूचकांक का प्रस्ताव करने का प्रयास किया गया है। इस सूचकांक की प्रभावकारिता को वास्तविक समय के आंकड़ों के साथ प्रदर्शित किया गया है। परिणामों से यह स्पष्ट होता है कि किसी जीनोटाइप के सापेक्ष प्रदर्शन के आधार पर गणना की गई गैर-पैरामीट्रिक उपाय दूसरों की तुलना में, व्यावहारिक रूप से या तो विविधता के साथ या एक आशाजनक लाइन के रूप में जारी करने के लिए स्थिर लाइनों के साथ आने के लिए अधिक व्यावहारिक रूप से आगामी फसल संकरण ट्रेल्स सार्थक हो सकते हैं। काम किए गए लक्षण ाों के महत्व का मापन उनके भविष्य के संकरण ट्रेल्स में प्रजनकों के लिए चयन मानदंड के रूप में भी हो सकता है। यह वैराइटी रिलीज प्रोग्राम में इस पद्धति का उपयोग करने का सुझाव दिया गया है और इसे फसल की किस्मों के मल्टी-लोकेशन ट्रेल (एमएलटी) आधारित रिलीज के लिए बढ़ाया जा सकता है।

टेक्नोलोजी टेक्स्ट और यूएसडीए सॉयल टैक्सोनॉमी ओन्टोलॉजी से ओन्टोलॉजी के विकास के लिए ओन्टोलॉजी एल्गोरिदम सीखना

चंदन कुमार देब[1], सुदीप मारवाहा[1] एवं आर.एन. पांडेय[2]

[1]भा.कृ.अ.प. - भारतीय कृषि सांख्यिकी अनुसंधान संस्थान, नई दिल्ली

[2]भा.कृ.अ.प. - भारतीय कृषि अनुसंधान संस्थान, नई दिल्ली

वेब आधारित सॉफ्टवेयर पारंपरिक वेब से एक कदम आगे वेब को बनाए रखने के लिए ओन्टोलॉजी का उपयोग करते हैं। ओन्टोलॉजी आधारित सॉफ्टवेयर आर्किटेक्चर प्लेटफॉर्म को मानव के साथ-साथ मशीन के लिए भी उपयुक्त बनाता है। ओन्टोलॉजी के माध्यम से असंरचित ज्ञान आसानी से संरचित एक में परिवर्तित हो जाता है। दास (2010) और दास एट अल द्वारा यूएसडीए मिट्टी वर्गीकरण के लिए मृदा वर्गीकरण स्वायत्तता विकसित की गई। (2012) भारत में उपलब्ध मिट्टी के आदेश केवल उप समूह स्तर के लिए उपलब्ध है। इस काम में विकसित नॉलेजबेस का उपयोग एन-टियर आर्किटेक्चर के साथ वेब आधारित सॉफ्टवेयर विकसित करने के लिए किया जाता है और ओन्टोलॉजी को परिवार और श्रृंखला स्तर तक बढ़ाया गया है। यह यूएसडीए मृदा वर्गीकरण के सभी बारह आदेशों को भी शामिल करता है और मिट्टी के राज्यवार श्रृंखला विवरण प्रदान करता है। सॉफ्टवेयर का खोज मॉड्यूल मिट्टी की वर्गीकरण की अनन्य खोज प्रदान करता है और संपादन मॉड्यूल ओन्टोलॉजी जानकारी की सुविधा को जोड़ने, हटाने और संपादित करने की सुविधा प्रदान करता है। इसके अतिरिक्त हमने टैक्नोनामिक ग्रंथों से स्वचालित ओन्टोलॉजी सीखने के लिए एक एल्गोरिथम विकसित किया है, जो मिट्टी के वर्गीकरण के केस स्टडी के साथ है।

BOOK REVIEW

Statistics for Agricultural Sciences

G. Nageswara Rao

Former Professor & Head, Acharya N.G. Ranga Agricultural University, Hyderabad, Telangana

Publisher: B.S. Publications, Hyderabad

by

Ramasubramanian V.

ICAR-Indian Agricultural Statistics Research Institute, New Delhi

---

The author has painstakingly written this book on the subject of statistics to cater to the needs of students who pursue their tertiary education in agricultural and allied sciences with supposedly having not much mathematical background during their schooling. This third edition of his work has some additions with respect to few portions on Econometrics and also a flavour of using SPSS software for data analysis as an added feature. The whole book has 21 chapters in all written under five major parts. While the chapter on Experimental Designs runs into over 120 pages out of the total 460 pages, some chapters are very small.

The first chapter gives an introduction to the subject of statistics. However, it would have been better if its usefulness in the field of agricultural sciences could also have been touched upon by means of bringing in few practical situations. The second chapter on collection of data and related portion dwells on the topic on expected lines. The recent elicitation of information and data via electronic mode of communication could have given a different dimension to the treatment. The third chapter on frequency distribution includes pictorial and graphical representation of data as well which usually can be seen along with classification and tabulation chapters in other books. The figures of multiple bar diagram and Pie diagram got shifted by one subsection even though fortunately they appear on pages that face such subsections. A noteworthy inclusion in this chapter is the discussion on Lorenz curve which brings novelty to the book followed by an illustration on fitting the same which is commendable. The fourth chapter on measures of location is dealt with in a routine way when discussion on situations under which, say, geometric and harmonic means, are more suitable could have enriched its content. The section on 'Sample' under the fifth chapter on measures of dispersion is misplaced or it could have been titled as Standard error. The sixth chapter on moments, skewness and kurtosis is written in a simple manner, with perhaps a minor typo in the relationship between the coefficient of skewness β_1 and its γ_1 counterpart. The numerical figures appearing in the whole book could have been right justified throughout for proper vertical alignment of decimals and unit places.

The chapter on Probability is written in a lucid manner which also includes some portion of binomial distribution which rightfully should belong to the chapter that follows. The chapter on binomial and Poisson distributions also includes practical exercises on their fitting to data. The ninth chapter on Normal distribution starts with a mathematical angle of relating binomial to normal distribution which may appeal to those who have more of a statistical bent of mind but may baffle many an intended reader. Here again fitting of a normal distribution has been included which completes the practical aspect in this series of fitting of distributions. The chapter on tests of hypotheses is written well but the insertions with regard to SPSS data analysis suffers from readability. The chapter on Chi-square distribution has some new topics such as Dandekar's method of correction of continuity (in addition to Yates' method) in case of 2x2 contingency table and also a subsection on Chi-square for testing linkage between genes. Having said that, the treatment of SPSS again in this chapter has been given as a lengthy procedure while there exist many easy way of doing the same thing via SPSS itself, say, for Chi-square test for a 2x2 table of association between two attributes. Twelfth chapter on correlation and regression is written in a good manner but for testing significance of correlation the author has started with Fisher's Z transformation by considering the situation when the population correlation coefficient is not zero. Later, under rank correlation, the usual t statistic expression could be seen. This way of writing departs from the usual practice.

In the chapter on Multiple regression and correlation, the solving of normal equations by matrix method may be beyond the level of the intended students, rather a short cut method could have been given. The next chapter on D^2 statistic and discriminant functions again should have been a part of the multivariate statistical methods in Chapter 20 and its discussion as a separate chapter here is surprising. Here again the mention of 'pivotal condensation method' etc. for computing D^2 statistic should be a wee bit tough for the students with not so good mathematical background. Chapter 15 on Probit analysis is praiseworthy with both biological and economic data used for explaining its utility. The chapter on Experimental designs is an elaborate one which apart from including the common designs like CRD, RBD and LSD also deals with much special type of designs like split-split designs, Lattice designs and cross over designs. The explanation on interaction effect by means of diagrams stands apart. The estimation of mixed models using Henderson's method also finds its place under this chapter which may be quite useful to understand it better. The inclusion of path coefficient analysis inside the experimental designs chapter may not sometimes be seen by the reader even though a worked out example is given.

The chapter on Sampling is also exhaustive and the author has tried to tabulate the expressions of estimates of sample mean/ total etc. and their population counterparts. A section on tolerance in testing of seeds is given under this chapter of sampling which again seems to be out of place. Under the chapter on Economic statistics, the topics ranging from elements of time series analysis, index numbers, fitting of growth curves are given. While the effort made is appreciable, in this era, direct fitting of non-linear growth curves is warranted rather than suggesting linearization by logarithmic transformation before fitting them which are at the most approximations if not inappropriate. The chapters on Non-parametric statistics and multivariate statistical methods that follow are rich in numerical examples in the field of agriculture and hence may directly strike a chord with the student audience for which this book has been written. The last chapter on Econometrics (newly added in this edition) contains material with more mathematical rigour which could have been toned down for easy comprehension by the expected readers.

Overall, the book has been written in a comprehensive manner with a well meaning intention of usefulness to the students. It is hoped that the book brought out will serve as an excellent source of knowledge to the students and also will help them in applying appropriate statistical techniques to their agriculture related real life data sets. I congratulate the author for bringing out this valuable book for the benefit of students and researchers in agricultural and allied sciences.

LEPING LI, Biostatistics Branch, NIEHS, Alexander Dr. RTP, NC 27709, USA. E-mail: li3@niehs.nih.gov
RPS MALIK, E-407, Greater Kailash, Part-II, New Delhi - 110048, India. E-mail: rps_malik@yahoo.com
YING MIAO, University of Tsukuba, Tsukuba 305-8573, Japan. E-mail: miao@sk.tsukuba.ac.jp
JP MORGAN, Virginia Tech Blacksburg, VA 24061-0439, USA. E-mail: jpmorgan@vt.edu
RAHUL MUKERJEE, Indian Institute of Management, Calcutta, Kolkata - 700104, India.
E-mail: rmuk@iimcal.ac.in
BALGOBIN NANDRAM, Department of Mathematical Sciences, Worcester Polytechnic Institute,
100 Institute Road, Worcester, MA 01609-2280, USA. Email: balnan@wpi.edu
NAM-KY NGUYEN, Vietnam National University, Building C, Hacinco Student Village Nhan Chinh,
Thanh Xuan Hanoi, Vietnam. E-mail: namnk@isvnu.vn
SHYAMAL D PEDDADA, Biostatistics Branch, NIEHS, Alexander Dr. RTP, NC 27709, USA.
E-mail: peddada@niehs.nih.gov
PRAJNESHU, ICAR-IASRI, New Delhi - 110012, India. E-mail: prajneshu@yahoo.co.in
MONICA PRATESI, University of Pisa, Via Ridolfi, 10, 56124-Pisa, Italy. E-mail: m.pratesi@ec.unipi.it
CRISTINA RUEDA SABATER, University of Valladolid, Valladolid, Spain. E-mail: crueda@eio.uva.es
NICOLA SALVATI, University of Pisa, Italy. E-mail: salvati@ec.unipi.it
SD SAMANTARAY, GB Pant University of Agriculture & Technology, Pantnagar - 263145, India.
E-mail: sdsray@gmail.com
DEBARAJ SEN, Concordia University, Montreal. Canada. E-mail: debaraj.sen@concordia.ca
SD SHARMA, D-502, SF, Presidency Floor, ARDEE City, Sector 52, Gurgaon - 122011. India.
E-mail: sdsharmaus@gmail.com
VK SHARMA, C-105, Rudra CGHS Ltd., Plot # 12, Sector, Sector 6, Dwarka, New Delhi - 110075, India.
E-mail: vksharma_10@yahoo.co.in
MURARI SINGH, ICARDA, Beirut, Lebanon. E-mail: mandrsingh2010@gmail.com
RANDHIR SINGH, A-504, Arvind Apartments, Sector 19B, Dwarka, New Delhi - 110075, India.
E-mail: rsdahiya7@yahoo.co.in
BIKAS K SINHA, Indian Statistical Institute, Kolkata - 700108, India. E-mail: bikassinha1946@gmail.com
BVS SISODIA, H-1/799, Janakipuram, Lucknow - 226 021, India. E-mail: bvssisodia@gmail.com
TUMULESH SOLANKY, University of New Orleans, New Orleans, LA 70148, USA.
E-mail: tsolanky@uno.edu
AK SRIVASTAVA, B-25/G-1, Dilshad Garden, Delhi - 110095, India. E-mail: arunsrivast@gmail.com
NIKOS TZAVIDIS, University of Southampton, UK. E-mail: n.tzavidis@soton.ac.uk