International Journal of Mathematical Sciences and Optimization: Theory and Applications
12(1), 2026, Pages 129 - 142
https://doi.org/10.5281/zenodo.20467453

Ratio Estimation of Population Proportion With Optimality in Presence of Non-Response

M. Cherotich [1,2*], D. A. Alilah [2], E. Okuto [2]

1. Department of Mathematics, Masinde Muliro University of Science and Technology Kakamega-Nairobi, Kenya.
2. Department of Mathematics, Jaramogi Oginga Odinga University of Science and Technology Siaya -Nairobi, Kenya.
Email: [1]chemoigutmercy@gmail.com, [2]erickokuto@gmail.com, [2]aliladavid2010@gmail.com
* Corresponding author: chemoigutmercy@gmail.com

Article Info
Received: 24 August 2024 Revised: 10 February 2025
Accepted: 30 March 2025 Available online: 20 May 2026

Abstract

Studies on the estimation of population proportions have been around at improving the accuracy and efficiency of survey designs. A number of studies have been carried out on estimation of population proportion without non-response under optimality, however,non-response remains a significant challenge , often introducing bias and reducing reliablity. This study examined the impact of non-response on sample size, bias, variance and relative efficiency using the ratio estimation. Simulations study has been performed using R-Sofware to compare empirical estimator of the population proportion. Results indicated that an increase in the non-response rate leads to a larger sample size, increased MSE and reduced bias and variance. Graphical analysis confirmed that the MSE increased the sample size, highlighting the limitations of large samples in the presence of non-response. From the results it was confirmed that the ratio estimators were reliable method for proportion estimation provided non-response adjustments and optimal sample allocation are implemented.

Keywords: Ratio estimation, Optimality, Population proportion, Study variable and auxiliary variable.
MSC2010: 35Q20.

1 Introduction

A proportion is a ratio of the values of subset to the values of a set. Estimation of proportion has been discussed from the perspective of various sampling plans ([1],([2]). According to [3] he presented estimators for a proportion using the logistic regression estimator. The study demonstrated that logistic models efficiently facilitated a better modelling of the survey data. Pierre [3] estimated proportion using various sampling plans that were Bernoulli and Stratified sampling designs. The empirical results showed that applying four sampling plans to real data set improved the efficiency

http://ijmso.unilag.edu.ng/article

of estimators through the appropiate use of auxiliary variables and model selection. However, the study did not focus on obtaining the optimal value of the estimators, leaving a gap in achieving the best possible estimator performance.

Martinez [4] explored the problem of estimating the population proportion for categorical variable using the calibration framework. The study investigated auxiliary information and theoreical properties, leading to the defination of a new class of calibration-base optimal estimators. A simulation study evaluated the performance of the proposed calibration estimator via empirical relative bias and emprical relative efficiency ,yielding favorable results. However, the study did not establish an optimal value for the estimator, highlighting a gap in determining the best estimator performance..

Alvarez [5] estimated the population proportion in the presence of missing data and using auxiliary information at the estimation stage. The sudy proposed a general class of estimators that efficiently utilized availbale information and the theorical properties of these estimators were analyzed and it gave them the best estimator. Despite these contributions , the study did not explicity address optimality in the presence of non-response, leaving room for further research on efficiency improvements.

Adeyemi [6] did a study on properties of exponential pareto distribution, this study investigated the stochastic ordering properties, moments of order statistics and some distributional properties. The computational results from higher moments of order statistics provides some characterizations for the exponential Pareto distribution, variability ordering was obtained for equal and unequal sample sizes and the results strenghtened some stochastic ordering existing in the literature. The study did not focus on population proportion, hence the research gap.

In survey sampling, usually it is assumed that all the observations are correctly measured for the characteristic under study. But in practice, this assumption is not met for a variety of reasons, such as non-response which may occur due to the refusal of respondents to give some information. Usually measurement and non-response error are computed separately using a known auxiliary or additional information. In reality, both measurement and non-response error occur simultaneously in survey sampling. Most, information is not obtained from all the units during surveys resulting to non-response errror as a common problem during sampling.

In survey sampling, the estimation of population mean of a variable of interest in the presence of non-response, when the auxiliary information available was widely considered by [7]. Also [8] suggested an improved class of estimators of finite population mean in the presence of measurement and non-response error under stratified random sampling. By use of simulation study and real life data sets it was confirmed that the proposed class of estimators performed better than the global estimators. Nevertheless, neither study explicitly addressed the optimization of thes estimators under non-response conditions, which is a key research gap.

Garg [9] proposed an estimator on calibration estimation of population proportion in probability proportional to size sampling in presence of non-response. They addressed the problem of estimation of finite population under the probability proportional to size(pps) sampling techniques,when the complete information is unavailable due to presence of non-response. They developed calibrated estimators of population proportion under PPS in the presence of non-response based on the availability of auxiliary information. The study did not focus on achieving optimality, leaving a gap in identifying the best calibration estimator under non -response.

On issue estimation, Okafor [10] did a study on the maximum likelihood estimation of hidden markov model where estimation was done for infectious disease progression. In the study a hidden markov model was formulated to estimate the rate of infectious disease. This study focused on estimation using markov chain but not estimation using population proportion and hence the gap in the literature.

Kumar [11] proposed a class of estimators for the population mean under different sampling designs in presence of non-response. In their proposition, non-response was computed separately using auxiliary variables. Anekeya [12] proposed estimation of the domain mean using double sampling with a non-linear cost function in the presence of non-response. They proposed estimation of the domain mean using auxiliary information and the mean square errors of the proposed estimators was obtained. Optimal stratum sample sizes for the given set of non-linear cost function was also

International Journal of Mathematical Sciences and Optimization: Theory and Applications
12(1), 2026, Pages 129 - 142
https://doi.org/10.5281/zenodo.20467453

developed. The proposed class of estimators of sample sizes decreased as the sub-sampling fraction together with the number of auxiliary variables increased. However, these studies did not address the proble of achieving an optimal estimator that balances efficiency and accuracy in non-response. The use of study variables and auxiliary variables in ratio estimation is essential for achieving optimal estimates of population proportions, particularly in the presence of non-response. However, the accuracy of ratio estimation can be compromised in the presence of non response, which is a common problem in surveys. Therefore, developing methods to improve the accuracy of ratio estimation in the presence of non response is crucial for obtaining accurate estimates. Secondly, non-response introduces bias into survey estimates and reduces the efficiency of estimation. This can have significant implications for policy making, as inaccurate estimates can lead to incorrect decisions. Therefore, it was important to develop methods to address non response bias and to improve the efficiency of estimation. This study aims to bridge the identified research gaps by exploring ratio estimation of population proportion with optimality in the presence of non-response, an important research area with the potential to enhance the accuracy and efficiency of survey estimates and advance our understanding of social phenomena.

2 Materials and Methods

Ratio estimation of the population proportion using simple random sampling was developed and sub-sampling of the study variable and the auxiliary variable was employed, further the effect of the cost function on the optimal sample size was also considered.

2.1 Formulation of Population Proportion Estimator

Define a population of size N_p, divide the population size N_p into two disjoint groups of N_{p1} comprising of the responding group and $N_{p2}$of the non responding group. Further define y_{pi} and x_{pi} to be the study and the auxiliary attribute for the i^{th} unit of the population proportion such that $i = 1, 2, ..., N_p$.

Define the study variable y_{pi} as

$$y_{pi} = \begin{matrix} 1, & \text{If ith unit belongs to N\_p1} \\ 0, & \text{If ith unit belongs to N\_p2} \end{matrix}$$

Let

$$P_y = \sum_{i=1}^{N_p} \frac{y_{pi}}{N_p}$$

be the proportion of the population possessing attribute Y_p such that $\sum_{i=1}^{N} y_{pi} = Z_p$. This can further be expressed as

$$P_y = \frac{Z_p}{N_p}$$

;$1 - P_y = Q_y = 1 - \frac{Z_p}{N_p}$ Define

$$P_{y1} = \frac{\sum_{i=1}^{N_1 y_{pi}}}{N_{p1}} = \frac{Z_{p1}}{N_{p1}}$$

to be the proportion of the units responding and possessing attribute Y_p and

$$P_{y2} = \frac{\sum_{i=1}{}^{N_{p2} y_{pi}}}{N_{p2}} = \frac{Z_{p2}}{N_{p2}}$$

International Journal of Mathematical Sciences and Optimization: Theory and Applications
12(1), 2026, Pages 129 - 142
https://doi.org/10.5281/zenodo.20467453

be the proportion of the non responding population possessing attribute Y_p.

Population of size n_p is drawn from the population of size N_p using simple random sampling without replacement method (SRSWOR) from which attribute of size n_{p1} respond and those n_{p2} do not respond ,such that ,

$$n_{p1} + n_{p2} = n_p$$

Sub-sampling is further carried out and a sample of size r_p from n_{p2} non responding units and

$$m_p r_p = n_{p2}; r_p = \frac{n_{p2}}{m_p}, m_p > 1$$

The estimation of the population proportion of the study variable is defined as follows using the [13] technique.

$$\hat{P}_y = \frac{n_{p1}}{n_p} P_{y1} + \frac{n_{p2}}{n_p} P^*_{yrp} \tag{2.1}$$

such that
$P_{y1} = \frac{\sum_{i=1}{}^{n_{p1}} y_{pi}}{n_{p1}}$ is the proportion of the sample possessing the attribute y from n_1 non responding sample of the population of size n_1. $p^*_{yrp} = \frac{\sum_{i=1}{}^{r_p} y_{pi}}{r_p}$ is the proportion of the unit possessing the attribute y_p from the r_p sub sub-sampled sample proportion from n_{p2} non respondents.

Similarly the estimate for the population proportion of the auxiliary variable is given by

$$\hat{P}_x = \frac{n_{p1}}{n_p} p_{x1} + \frac{n_{p2}}{n_p} P^*_{xrp} \tag{2.2}$$

where P_{x1} and P^*_{x2} are the sample proportion from the population that respond and non respondent respectively. Define the weights of the responding units as follows

$$W_{p1} = \frac{N_{p1}}{N_p}$$

and

$$W_{p2} = \frac{N_{p2}}{N_p}$$

respectively. The estimated weights are further defined as

$$\hat{W}_{p1} = w_{p1} = \frac{n_{p1}}{n_p}$$

and

$$\hat{W}_{p2} = w_{p2} = \frac{n_{p2}}{n_p}$$

respectively. In estimating the population proportion in the presence of non response ratio estimation is used

$$\hat{P}_{Ry} = \frac{p_y}{p_x} . P_x = R_p P_x \tag{2.3}$$

with the assumption that

$$E[\hat{P}_y] = P_y$$

$$E[\hat{P}_x] = P_x$$

2.2 The Unbiased Estimator of the Population Proportion.

2.2.1 Proposition 1

The unbiased estimator of the population proportion of the study and auxiliary variables P_y and P_x for the variables Y and X respectively are given by

$$p_y = \frac{n_{p1}p_{y1} + n_{p2}p_{yrp}}{n_p} \tag{2.4}$$

and

$$p_x = \frac{n_{p1}p_{x1} + n_{p2}p_{xrp}}{n_p} \tag{2.5}$$

Proof
The equation (4) is given by

$$\hat{P}_y = \frac{n_{p1}}{n_p}p_{y1} + \frac{n_{p2}}{n_p}p_{yrp}$$

finding the expectation on both sides we have

$$E[\hat{P}_y] = E[\frac{n_{p1}}{n_p}p_{y1} + \frac{n_{p2}}{n_p}p_{yrp}]$$

but $w_{p1} = \frac{n_{p1}}{n_p}$ and $w_{p2} = \frac{n_{p2}}{n_p}$ then

$$E_1E_2[w_{p_1}p_{y1} + w_{p2}p_{yrp}]$$

expanding the above equation we have

$$E_1[w_{p1}p_{y1} + w_{p2}p_{yp}]$$

$$W_{p1}P_{y1} + W_{p2}P_{y2}$$

$$P_y \tag{2.6}$$

Similarly to the equation (5).

2.3 Bias of the population ratio estimator $\hat{P}_{Ry} = \frac{p_y}{p_x}p_x$

2.3.1 Proposition 2

The bias of the population proportion ratio estimator $\hat{P}_{Ry} = \frac{p_y}{p_x}p_x$ is given equation (8)
Proof

$$Bias(\hat{P}_{Ry}) = P_y[E(\epsilon_{p0}) - E(\epsilon_{p1}) + E(\epsilon_{p1}^2) - E(\epsilon_{p0}\epsilon_{p1})]$$

since $E(\epsilon_{p0}) = E(\epsilon_{p1}) = 0$
Therefore

$$Bias(\hat{P}_y) = P_y[E(\epsilon_{p1}^2) - E(\epsilon_{p0}\epsilon_{p1})]$$

$$P_y[\frac{N_p - n_p}{N_p - 1}\phi_x + (\frac{N_{p2}}{N_{p2} - 1})(\frac{m_p - 1}{n_p})W_{p2}\phi_{x2}\frac{P_{x2}}{P_x} - (\frac{N_p - n_p}{N_p - 1})Q_yQ_x] \tag{2.7}$$

But $\phi_x = \frac{Q_x}{P_x}$, $\phi_{x2} = \frac{Q_{x2}}{P_x}$
Therefore,

$$P_y[(\frac{N_p - n_p}{N_p - 1})\phi_x + (\frac{N_p}{N_{p2} - 1})(\frac{m_p - 1}{n_p})W_{p2}\phi_{x2}\frac{P_{x2}}{P_x} - (\frac{N_p - n_p}{N_p - 1})Q_yQ_x]$$

$$P_y[(\frac{N_p - n_p}{N_p - 1})\frac{Q_x}{P_x} + (\frac{N_p}{N_{p2} - 1})(\frac{m_p - 1}{n_p})W_{p2}\frac{Q_{x2}}{P_x}\frac{P_{x2}}{P_x} - (\frac{N_p - n_p}{N_p - 1})Q_yQ_x]$$

$$= (\frac{N_p - n_p}{N_p - 1})\frac{P_y}{P_x}Q_x + (\frac{N_p}{N_{p2} - 1})(\frac{m_p - 1}{n_p})W_{p2}\frac{P_y}{P_x}Q_{x2}\frac{P_{x2}}{P_x} - (\frac{N_{p-n_p}}{N_p - 1})P_yQ_yQ_x$$

$$= (\frac{N_p - n_p}{N_p - 1})R_pQ_x + (\frac{N_p}{N_{p2} - 1})(\frac{m_p - 1}{n_p})W_{p2}R_pQ_{x2}\frac{P_{x2}}{P_x} - (\frac{N_{p-n_p}}{N_p - 1})P_yQ_yQ_x$$

$$= (\frac{N_p - n_p}{N_p - 1})R_pQ_x + (\frac{N_p}{N_{p2} - 1})(\frac{m_p - 1}{n_p})R_pW_{p2}\phi_{x2}P_{x2} - (\frac{N_p - n_p}{N_p - 1})P_yQ_yQ_x$$

$$Bias(\hat{P}_{Ry}) = (\frac{N_p - n_p}{N_p - 1})R_pQ_x + (\frac{N_p}{N_{p2} - 1})(\frac{m_p - 1}{n_p})R_pW_{p2}\phi_{x2}P_{x2} - (\frac{N_p - n_p}{N_p - 1})P_yQ_yQ_x \quad (2.8)$$

where $R_p = \frac{P_y}{P_x}$

2.4 MSE of the population proportion ratio estimator

2.4.1 Proposition 3

The mean square error (MSE) of the proportion ratio estimate $\hat{P}_{Ry} = \frac{p_y}{p_x}P_x$ is given by equation (9)
Proof

$$MSE(\hat{P}_{Ry}) = E[\hat{P}_{Ry} - P_{Ry}]^2$$

$$= E[\frac{P_y}{P_x} - P_y]^2$$

but $p_y = P_y(\epsilon_{p0} + 1)$,$p_x = P_x(\epsilon_{p1} + 1)$ Substituting the above in the expression it becomes

$$= E[\frac{p_y(\epsilon_{p0} + 1)}{(\epsilon_{p1} + 1)} - P_y]^2$$

International Journal of Mathematical Sciences and Optimization: Theory and Applications
12(1), 2026, Pages 129 - 142
https://doi.org/10.5281/zenodo.20467453

$$= E[P_y(\frac{\epsilon_{p0}+1}{\epsilon_{p1}+1}) - 1]^2$$

$$= P_y^2 E[\frac{(\epsilon_{p0}+1)-(\epsilon_{p1}+1)}{\epsilon_{p1}+1}]^2$$

$$= P_y^2 E\frac{(\epsilon_{p0}+1)-(\epsilon_{p1}-1)}{\epsilon_{p1}+1}]^2$$

$$= P_y^2 E[\frac{(\epsilon_{p0}-\epsilon_{p1})}{(\epsilon_{p1}+1)}]^2$$

$$= P_y^2 E[(\epsilon_{p0}-\epsilon_{p1})(\epsilon_{p1}+1)^{-1}]^2$$

$$= P_y^2[(\epsilon_{p0}-\epsilon_{p1})(1+\epsilon_{p1})^{-1}]^2$$

By Taylor series and approximation the expression becomes

$$= P_y^2 E[(\epsilon_{p0}-\epsilon_{p1})(1-\epsilon_{p1}+\epsilon_1^2+...)]^2$$

$$= P_y^2 E[\epsilon_{p0}-\epsilon_{p0}\epsilon_{p1}+\epsilon_{p0}\epsilon_{p1}^2-\epsilon_{p1}+ \epsilon_{p1}^2-\epsilon_{p1}^3+...]^2$$

$$= P_y^2[\epsilon_{p0}-\epsilon_{p1}]^2$$

$$= P_y^2 E[(\epsilon_{p0}^2)+(\epsilon_{p1})^2-2\epsilon_{p0}\epsilon_{p1}]$$

$$= P_y^2 E[\epsilon_{p0}^2+\epsilon_{p1}^2-2\epsilon_{p0}\epsilon_{p1}]$$

$$= P_y^2[(\epsilon_{p0}^2)+E(\epsilon_{p1})^2-2E(\epsilon_{p0}\epsilon_{p1})]$$

Using ratios the expression is then simplified to

$$= P_y^2[(\frac{N_p-n_p}{N_p-1})\frac{Q_y}{P\ _y}+(\frac{N_{p2}}{N_{p2}-1})(\frac{m_p-1}{n_p})\frac{P_{y2}Q_{y2}W_{p2}}{P_{\frac{2}{y}}}+ (\frac{N_p-n_p}{N_p-1})\frac{Q_x}{p_x}+ (\frac{N_{p2}}{N_p-1})(\frac{m_p-1}{n_p})\frac{P_{x2}Q_{x2}}{P_{x^2}}W_{p2}-2[(\frac{N_p-n_p}{N_p-1})Q_yQ_x]]$$

$$= (\frac{N_p-n_p}{N_p-1})Q_yP_y+(\frac{N_{p2}}{N_{p2}-1})(\frac{m_p-1}{n_p})P_{y2}Q_{y2}W_{p2}+ (\frac{N_p-n_p}{N_p-1})Q_xP_y\frac{P_y}{P_x}+(\frac{N_{p2}}{N_p-1})(\frac{m_p-1}{n_p})R_p^2P_{x2}Q_{x2}W_{p2}- 2(\frac{N_p-n_p}{N_p-1})P_y^2Q_yQ_x$$

IJMSO
International Journal of Mathematical Sciences and Optimization: Theory and Applications
12(1), 2026, Pages 129 - 142
https://doi.org/10.5281/zenodo.20467453

By simplification we have

$$= (\frac{N_p - n_p}{N_p - 1})Q_y P_y + (\frac{N_p - n_p}{N_p - 1})R_p Q_x P_y +$$
$$(\frac{N_{p2}}{N_{p2} - 1})(\frac{m_p - 1}{n_p})P_{y2}Q_{y2}W_{p2} + (\frac{N_{p2}}{N_p - 1})(\frac{m_p - 1}{n_p})R_p^2 P_{x2}Q_{x2}W_{p2} -$$
$$2(\frac{N_p - n_p}{N_p - 1})Q_y Q_x P_y^2$$

$$= (\frac{N_p - n_p}{N_p - 1})Q_y P_y + R_p Q_x P_y - 2P_y^2 Q_y Q_x + (\frac{N_{p2}}{N_{p2} - 1})(\frac{m_p - 1}{n_p})P_{y2}Q_{y2}W_{p2} + R_p^2 P_{x2}Q_{x2}W_{p2}$$

On factorization and the simplification the expression reduces to

$$= (\frac{N_p - n_p}{N_p - 1})P_y(Q_y + R_p Q_x - 2P_y Q_y Q_x + (\frac{N_{p2}}{N_{p2} - 1})(\frac{m_p - 1}{n_p})W_{p2}(P_{y2}Q_{y2} +$$
$$R_p^2 P_{x2}Q_{x2})$$

$$= (\frac{N_p - n_p}{N_p - 1})P_y \Omega_{p1} + \frac{N_{p2}(m_p - 1)}{n_p(N_{p2} - 1)}W_{p2}\Omega_{p2} \tag{2.9}$$

Where $\Omega_{p1} = Q_y + R_y Q_x - 2P_y Q_y Q_x$, $\Omega_{p2} = P_{y2}Q_{y2} + R_p^2 P_{x2}Q_{x2}$

2.5 Optimum allocation in estimate of population proportion.

In optimal allocation a sample size , the precision is minimized against a given cost or a finite cost is minimized against precision. In developing the optimal sample size define the cost function as

$$C_p = c_p n_p + c_{p1} n_{p1} + c_{p2} m_p \tag{2.10}$$

where C_p -is the overhead cost sample.
c_p - cost of identity sampling target population n_p
c_{p1} -cost of measuring a unit in the response sample n_{p1}.
c_{p2} - represents the cost of measuring amount observed under sub-sample m_p from n_{p2}.
In computation of the optimal allocation of the sample size the variance is minimized against a given cost or cost is minimized over a given variance. Define,

$$m_{p2} r_{p2} = n_p$$

$$n_{p1} = W_{p1} n_p$$

$$m_{p2} = \frac{n_p}{r_{p2}} \Longrightarrow m_{p2} = \frac{w_{p2} n_p}{r_{p2}} = \frac{n_{p2}}{r_{p2}}$$

$$C_p = c_p n_p + c_{p1} W_{p1} n_p + c_{p2} W_{p2} \frac{n_p}{r_{p2}}$$

$$C_{p=} c_p n_p + n_p(c_{p1} w_{p1} + c_{p2} \frac{W_{p2}}{r_{p2}})$$

International Journal of Mathematical Sciences and Optimization: Theory and Applications
12(1), 2026, Pages 129 - 142
https://doi.org/10.5281/zenodo.20467453

2.5.1 Proposition 4

The variance of the estimated population ratio estimator $\hat{P}_{Ry}$ is minimum for the specified cost C_p when

$$n_p = \sqrt{\frac{N_{p2}W_{p2}\Omega_{p2}(N_p-1)(1-m_{p2})}{P_y\Omega_{p1}}}$$

$$r_{p2} = \sqrt{\frac{N_{p2}W_{p2}\Omega_{p2}(N_p-1)(1-m_{p2})}{m_{p2}^2 p_y\Omega_{p1}}}$$

Proof

$$\Theta(w_p) = (\frac{N_p-n_p}{N_p-1})p_y\Omega_{p1} + \frac{N_{p2}(m_{p2}-1)}{n_p(N_{p2}-1)}W_{p2}\Omega_{p2} + \lambda[c_p n_p + n_p c_{p1}W_{p1} + \frac{c_{p2}n_p W_{p2}}{r_{p2}-n_p c_p}]$$

but

$$m_{p2} = \frac{n_p}{r_{p2}}$$

$$\Theta(w_p) = (\frac{N_p-n_p}{N_p-1})p_y\Omega_{p1} + \frac{N_{p2}(m_{p2}-1)}{n_p(N_{p2}-1)}W_{p2}\Omega_{p2} + \lambda[c_p n_p + n_p c_{p1}W_{p1} + c_{p2}m_{p2}W_{p2} - n_p c_p]$$

Taking partial derivative with respect to m_{p2} and it equates to zero.

$$\frac{\partial\Theta(W_p)}{\partial(m_{p2})} = \frac{N_{p2}}{n_{pN_{p2}-1}}W_{p2}\Omega_{p2} + \lambda c_{p2}W_{p2} = 0$$

$$\frac{N_{p2}W_{p2}\Omega_{p2}}{n_p(N_{p2}-1)} = -\lambda c_{p2}W_{p2}$$

$$\frac{N_{p2}\Omega_{p2}}{n_p(N_{p2}-1)} = -\lambda c_{p2}$$

$$\frac{N_{p2}\Omega_{p2}}{n_p} = -\lambda c_{p2}(N_{p2}-1)$$

$$\frac{N_{p2}\Omega_{p2}}{n_p} = \lambda c_{p2} - \lambda c_{p2}N_{p2}$$

$$\lambda(c_{p2} - c_{p2}N_{p2}) = \frac{N_{p2\Omega_{p2}}}{n_p}$$

$$\lambda c_{p2}(1-N_{p2}) = \frac{N_{p2}\Omega_{p2}}{n_p}$$

$$\lambda = \frac{N_{p2}\Omega_{p2}}{n_p c_{p2}(1-N_{p2})}$$

$$n_p = \frac{N_p\Omega_{p2}}{\lambda c_{p2}(1-N_{p2})} \tag{2.11}$$

$$\Phi(W_p) = (\frac{N_p - n_p}{N_p - 1})p_y\Omega_{p1} + \frac{N_{p2}(m_{p2}-1)}{n_p(N_{p2}-1)}W_{p2}\Omega_{p2} + \lambda[c_p n_p + n_p c_{p1}W_{p1} - n_p c_{p1}]$$

But

$$n_p = m_{p2}r_{p2}$$

$$\Phi(W_p) = (\frac{N_p - n_p}{N_p - 1})p_y\Omega_{p1} + \frac{N_{p2}(m_{p2}-1)}{n_p(N_{p2}-1)}w_{p2}\Omega_{p2} + \lambda[m_{p2}r_{p2}[c_{p1} + c_{p1}w_{p1} - c_p] + c_{p2}m_{p2}] \tag{2.12}$$

Next the partial derivative with respect to n_p is obtained by

$$\frac{\partial\phi(W_p)}{\partial n_p} = -\frac{p_y\Omega_{p1}}{N_p - 1} - \frac{N_{p2}(m_{p2}-1)}{n_p^2(N_{p2}-1)}W_{p2}\Omega_{p2} = 0$$

$$\frac{N_{p2}(m_{p2}-1)n_p^2(N_{p2}-1)}{W}{}_{p2}\Omega_{p2} = -\frac{p_y\Omega_{p1}N_p - 1}{}$$

$$= N_{p2}(m_{p2}-1)W_{p2}\Omega_{p2}(N_p - 1) = -p_y\Omega_{p1}n_p^2$$

$$= (N_{p2}m_{p2} - N_{p2})W_{p2}\Omega_{p2}(N_p - 1) = -p_y\Omega_{p1}n_p^2$$

$$= N_{p2}m_{p2}W_{p2}\Omega_{p2}(N_p - 1) - N_{p2}W_{p2}\Omega_{p2}(N_p - 1)$$

$$= -p_y\Omega_{p1}n_p^2 \tag{2.13}$$

$$N_{p2}m_{p2}W_{p2}\Omega_{p2}(N_p - 1) - N_{p2}W_{p2}\Omega_{p2}(N_p - 1) = -p_y\Omega_p n_p^2$$

$$N_{p2}m_{p2}W_{p2}\Omega_{p2}(N_p - 1) + p_y\Omega_{p1}n_p^2 = N_{p2}W_{p2}\Omega_{p2}(N_p - 1)$$

$$= p_y\Omega_{p1}n_{p^2} = N_{p2}W_{p2}\Omega_{p2}(N_p - 1) - N_{p2}m_{p2}W_{p2}\Omega_{p2}(N_p - 1)$$

$$p_y\Omega_{p1}n_p^2 = N_{p2}W_{p2}\Omega_{p2}(N_p - 1)[1 - m_{p2}]$$

$$n_p^2 = \frac{N_{p2}W_{p2}\Omega_{p2}(N_p - 1)[1 - m_{p2}]}{p_y\Omega_{p1}}$$

$$n_p = \sqrt{\frac{N_{p2}W_{p2}\Omega_{p2}(N_p-1)[1-m_{p2}]}{p_y\Omega_{p1}}} \tag{2.14}$$

but

$$n_p = m_{p2}r_{p2}$$

$$r_{p2} = \frac{n_p}{m_{p2}}$$

$$r_{p2} = \sqrt{\frac{N_{p2}W_{p2}\Omega_{p2}(N_p-1)(1-m_{p2})}{m_{p2}^2 p_y\Omega_{p1}}} \tag{2.15}$$

By determining the most efficient allocation of resources to different strata or sub-population, researchers can maximize the precision of their estimations while considering non-response issues.

3 Results and Discussion

3.1 Tabular representation of the results

Some of the results of the MSE , Bias and the optimal value to obtain a minimum cost function was obtained as at the table (1)below:

Table 1: Results of the MSE,Bias Variance and Relative efficiency under Optimal Values

Non-response	Sample size	MSE	BIAS	Variance	Relative Efficiency
1.25	334.2394	29.61859	0.2276875	0.2405688	0.8122223
1.50	472.6859	41.7345	0.1806078	0.1906567	0.4568001
1.75	578.9197	51.03659	0.1444822	0.1523579	0.2985268
2.0	668.4789	58.8761	0.1140269	0.1200704	0.2039375
2.25	747.2821	65.78287	0.08719529	0.09162464	0.1392834
2.50	818.7161	72.02706	0.06293763	0.0659067	0.09150404

3.2 Graphical representation of the MSE and Sample size and MSE and Non-response rate

https://doi.org/10.5281/zenodo.20467453

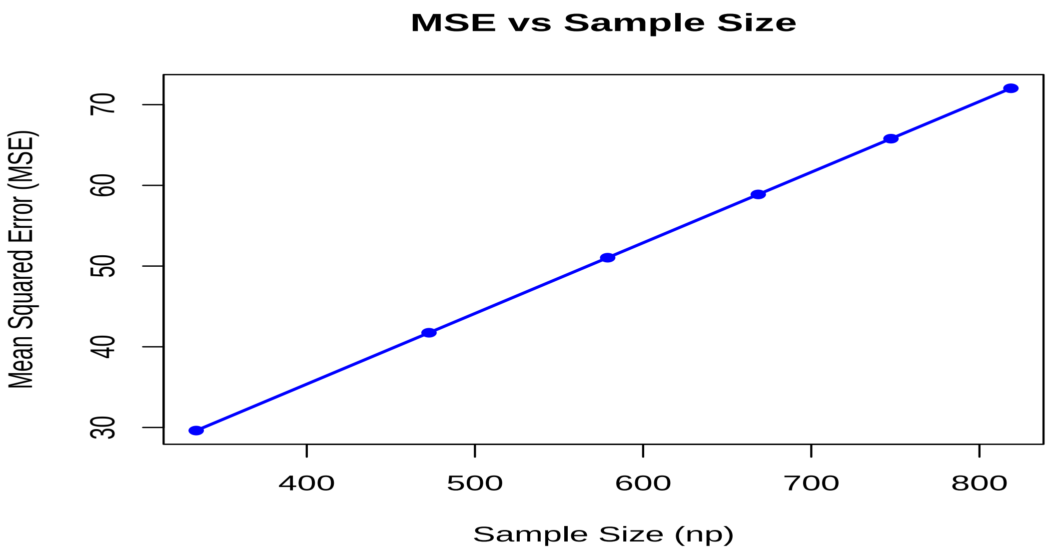

Figure 1: A line graph between the MSE and the Sample size

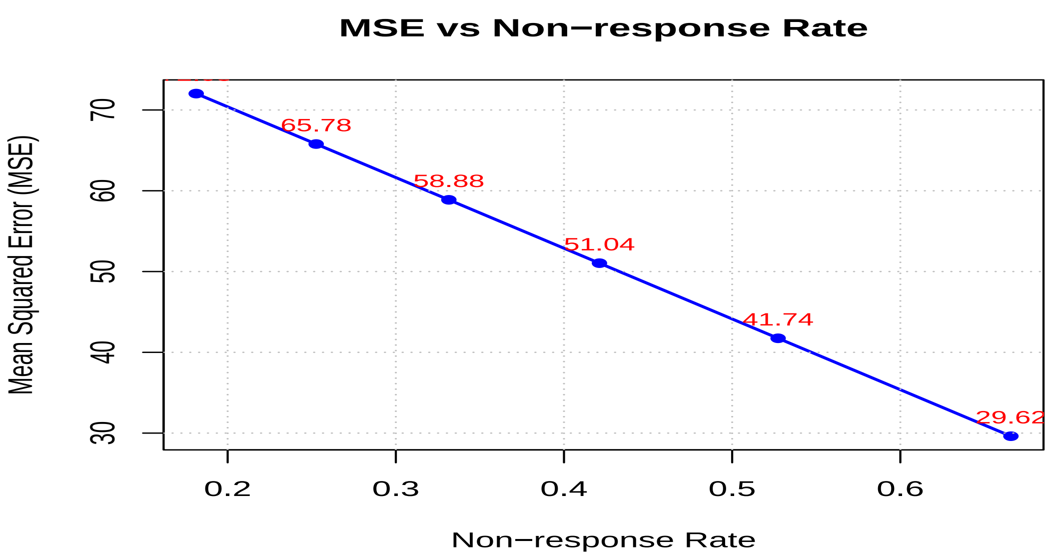

Figure 2: A line graph between the MSE and Non-response rate

3.3 Discussion

From the table above an increase in the non-response rate increased the sample size, with a subsequent increase in the MSE and a decrease in the Bias and variance. This implies that its sampling error of the population proportion is reduced for a large sample size.
An analysis of the efficiency of the computed estimates was determined using the Mean Squared Error (MSE), Bias, and the relative efficiencies. The Bias of the estimates was derived using Taylor's Series expansion. It was established that as the sample size n_p increased, the Bias of the population proportion decreased significantly. This implies that with an increase in the sample

size n_p, efficiency increases because the sampling mean becomes more normally distributed with reduced variability, leading to more accurate and reliable estimates of the population parameter.
It was observed in the graphical representation that the MSE increased significantly with an increase in the sample size, implying that the model error increased significantly. This illustrates that the estimates could no longer represent the entire population due to non-response bias. Cost analysis was used to determine the optimal sample size n_p, which increased significantly with an increase in the non-response rate.
The MSE decreased significantly with a significant decrease in the Non-reponse rate and typically an increse in the MSE increases the Non-response rate. Alvarez [5] did not address the optimality in the presense of non-reponse which was the identified in this study and from the results itmeans that that a lower non-response rate corresponds to a larger effective sample size, reducing variability and bias in the estimator as shown above.
It was also notable that there was a decrease in the variance (dispersion without auxiliary), which means that the estimates became more precise and reliable because they were based on more data points. An increase in the MSE (variation in the presence of auxiliary) meant that while there was reduced sampling variability (variance) due to the increased sample size, there was an introduced bias through the use of auxiliary information.

4 Conclusion

The main objective of the study was to estimate the population proportion with optimal allocation in the presence of non-response. The ratio estimation method is a valuable approach for estimating population proportions, even in the presence of non-response. This method utilized auxiliary variables that are correlated with the variable of interest, to produce an unbiased estimate of the population proportion. The ratio estimator was efficient when the auxiliary variables are highly correlated with the variable of interest and the response rate is high.
However, non-response may introduce bias into the ratio estimator, which can be reduced by applying the non-response adjustment technique or imputation methods. To ensure optimality, we chose the auxiliary variables carefully, considering their correlation with the variable of interest, and the sample size were selected to achieve the desired level of precision in the estimate. In summary, the ratio estimator was reliable and powerful for estimating population proportions since its bias was reduced and its variability reduced and it can be further improved in the presence of non-response by using non-response with optimality.

5 Acknowledgement

I acknowledge the immense support of my co-authors Dr. David Alila and Prof Eric Okuto for their continuous mentorship ,suggestion and support that led to the significant planning and execution of my research work. I would also like to thank all mathematics lecturers of Masinde Muliro University of science and Technology for their immense support during this study.

References

[1] Cochran, W.G.(1977). *Sampling Techniques* (3rd ed.).New York:John Wiley and Sons.

[2] Kish, L. (1965). *Survey sampling.* New York: John Wiley Sons.

[3] Pierre, D. (2003) Estimation of a Proportion with Survey Data,*Journal of Statistics Education,* 11(3), 2-9. https://doi.org/10.1080/10691898.2003.11910725

[4] Martinez, E. (2015) . Estimating population proportions by means of calibration estimators.*Statistics and Probability Letters.* 38, 267-293.

[5] Alvarez, E.(2013.) Estimating population proportions in the presence of missing data. Journal of Computational and Applied Mathematics. 237. 470–476. 10.1016/j.cam.2012.06.017.

[6] Adeyemi, A. (2023). Order Statistics and some of their Properties from the Exponential Pareto Distribution. International Journal of Mathematical Sciences and Optimization: Theory and Applications, 9(1), 15–32. Retrieved from https://www.ajol.info/index.php/ijmso/article/view/254325

[7] Khare, B. (2004). Estimation of finite population ratio using two-phase sampling Scheme in the presence of non-response. Aligarh Journal of Statistics. 24. 43-56.

[8] Zahid, E., Shabbir, J. (2018). Estimation of population mean in the presence of measurement error and non- response under stratified random sampling.*PLoS ONE,* 13(2), e0191572.https://doi.org/10.1371/journal.pone.019157

[9] Garg, N., Patel, A,. Pachori, M. (2024). Calibration Estimation of Population Proportion in Probability Proportional to Size Sampling in the Presence of Non-Response. *Indian Journal of Science and Technology* ,17(9), 841-851. https://doi.org/ 10.17485/IJST/v17i9.2938

[10] Okafor, R.U. (2024). Maximum Likelihood Estimation Of Hidden Markov Model: Application To Markers of Infectious Disease Progression. International Journal of Mathematical Sciences and Optimization: Theory and Applications, 10(3), 153–167. https://doi.org/10.5281/zenodo.13626109

[11] Kumar, S, (2016). Improved estimation of the population mean in the presence of non-response and measurement errors simultaneously.*Journal of Statistical Theoy and Practice.* 10(4),707-720.https://doi.org/10.1080/15598608.2016.1158642

[12] Alila D. (2018). Domain mean estimation using double sampling with non-linear cost function in the presence of non-response. *Science Journal of Applied Mathematics and Statistics.* 6(1), 28-42.

[13] Hansen,M.H., and Hurwitz, W.N.(1946). The problem of non-response in the sample surveys. *Journal of the American Statistical Association.* 41, 517-529.