An Efficient Poisson–Distributed Adaptive Cluster Sampling Model Using Randomized Response Strategy

4 authors, including:

Olumide Adesina
Redeemer's University
58 PUBLICATIONS   **137** CITATIONS

SEE PROFILE

Adedotun Adedayo
Olabisi Onabanjo University
42 PUBLICATIONS   **101** CITATIONS

SEE PROFILE

Gabriel Odekina
Covenant University Ota Ogun State, Nigeria
15 PUBLICATIONS   **37** CITATIONS

SEE PROFILE

An Efficient Poisson-Distributed Adaptive Cluster Sampling Model Using Randomized Response Strategy

Khalid Ul Islam Rather[1], Tanveer Ahmad Tarray[2], Olumide Sunday Adesina[3], Adedayo Funmi Adedotun[4*], Toluwalase Janet Akingbade[4], Onuche G. Odekina[4]

[1] Division of Statistics and Computer Science, Main Campus SKUAST-J, Jammu 180009, India
[2] Department of Mathematical Sciences, Islamic University of Science and Technology, Kashmir 192122, India
[3] Department of Mathematics and Statistics, Redeemer's University, Ede 232101, Nigeria
[4] Department of Mathematics, Covenant University, Ota 112212, Nigeria

Corresponding Author Email: adedayo.adedotun@covenantuniversity.edu.ng

ABSTRACT

The key innovation lies in the incorporation of an adaptive cluster sampling strategy and a randomized response model based on the Poisson distribution. This integration aims to overcome shortcomings inherent in conventional models, providing a more robust framework for research area. In this paper, an adaptive cluster sampling randomized response model with Poisson distribution using a randomized response strategy was proposed. The proposed cluster randomized response model has improved efficiency and a large gain in precision. Conditions were obtained under which the proposed model is more efficient than the existing models. To validate the effectiveness of our approach, numerical computations were conducted, offering concrete illustrations of the model's performance. The results underscore the significant gains in efficiency and precision achieved by the proposed adaptive cluster sampling randomized response model.

1. INTRODUCTION

In the realm of statistical sampling techniques, adaptive cluster sampling (ACS) has emerged as a powerful tool for studying rare and clustered populations. Its ability to enhance sampling efficiency in the presence of spatially aggregated units makes it particularly valuable for ecological and environmental studies, public health research, and various social sciences. However, one of the key challenges in implementing ACS is the potential bias introduced by non-responses and sensitive issues, which can undermine the reliability and validity of the collected data. To address these issues, this research proposes an innovative approach: the integration of a Poisson-distributed adaptive cluster sampling model with a randomized response strategy.

Adaptive cluster sampling is designed to capitalize on the natural clustering of certain populations. When an initial random sample includes units that meet a specified criterion, additional units in the neighborhood are included, continuing until no more units meet the criterion. This process leads to an adaptive expansion of the sample size in areas where the population is clustered, thereby increasing the efficiency of the survey. Despite its advantages, ACS can be vulnerable to biases and inaccuracies arising from non-responses, particularly when dealing with sensitive attributes or behaviors that respondents might be reluctant to disclose.

Adaptive cluster sampling is designed to capitalize on the natural clustering of certain populations. When an initial random sample includes units that meet a specified criterion,

additional units in the neighborhood are included, continuing until no more units meet the criterion. This process leads to an adaptive expansion of the sample size in areas where the population is clustered, thereby increasing the efficiency of the survey. Despite its advantages, ACS can be vulnerable to biases and inaccuracies arising from non-responses, particularly when dealing with sensitive attributes or behaviors that respondents might be reluctant to disclose.

Decades ago, research indicated that Warner [1] pioneered the development of the randomized response (RR) model. He applied this model to estimate proportions related to sensitive traits in humans, such as drug addiction and sexual orientation. However, the inclusion of unrelated questions in Warner's [1] model led Greenberg et al. [2] to introduce a modified approach known as the unrelated question randomized answer model. Since the introduction of Warner's randomized response model, it has garnered significant attention in the field of statistics. Various scholars, including authors of some studies [3-8], and others, have adopted and further developed the RR model.

A proportionate allocation-based stratified RR approach was first put forth by Hong et al. [9]. A decade later, Kim and Warde [10] introduced a stratified RR approach with optimal allocation. A two-stage RR model, as an application of the approach by Kim and Warde [10], was adopted in the study by Kim and Warde [11]. A stratified RR model and baseline from another source [12] were also referenced in their study. An updated Bayesian version of the Mangat model [13] was proposed by Kim et al. [14]. To address privacy issues and

extend the model to stratified sampling, Kim and Warde [15] proposed a mixed randomized response model. In a stratified sample using the Poisson distribution, Lee et al. [16] expanded upon the research conducted by Land et al. [17]. Current research on the topic is referenced in [18-23].

The integration of RR into the Poisson-distributed ACS model involves modifying the traditional ACS procedure to incorporate randomized response mechanisms at various stages of the sampling process. This dual approach aims to maintain the strengths of both methods while mitigating their individual limitations. The expected outcome is a sampling methodology that is not only efficient in terms of resource allocation and data collection but also robust against biases induced by non-responses and sensitive issues.

In the current study, we have proposed an adaptive cluster sampling randomized response model with Poisson distribution. This study aims to show that the proposed model is an improvement on the models existing in the literature.

This research endeavors to bridge the gap between adaptive sampling techniques and respondent confidentiality, presenting a novel Poisson-distributed adaptive cluster sampling model enhanced by a randomized response strategy. Through theoretical development and empirical validation, we aim to demonstrate the efficacy of this integrated approach in yielding high-quality, reliable data from clustered and sensitive populations. This innovative methodology holds promise for advancing the field of statistical sampling and improving the outcomes of various research endeavors reliant on accurate and unbiased data. Applications of this methodology span a wide range of fields, including epidemiology, where it can be used to study the prevalence of rare diseases; environmental science, for assessing the distribution of endangered species; and social sciences, for investigating sensitive behaviors and attitudes. By enhancing the integrity of data collection in these areas, the proposed model can contribute to more informed decision-making and policy development.

It is generally known that for adaptive cluster sampling design, if a given population is partitioned into N primary sample units containing $N_1, N_2, \ldots, N_n$ units. The procedure for two-stage sampling is done by (i) selecting a sample of size n from N using simple random sample (SRS) approach. (ii) In each selected primary sample units, a large sample m_i unit is selected from the N_i units. The remaining primary sample units are selected adaptively based on the condition used in estimating parameter in the primary sample units. The interval, C within the interest parameter's range, shall provide the requirement for the further collection of the neighbouring units. We take into account all the units on the list that are seen in the structure due to the first pick of unit i. Such a group, which could combine several neighbourhoods, is known as a cluster when it appears in a survey. A network of primary sample units exists within such a cluster and has the property that choosing one of the network's main sample units would result in the inclusion of all other primary sample units in the network's sample. When the unrelated rare attribute is known with the probability of a "Yes" response is given:

$$\phi_{i0} = [P_{1i}\beta_{if} + P_{2i}\beta_{iu}][1 + P_{3i}k_i/(k_i - 1)] \qquad (1)$$

where, $v_{i0} = [P_{1i}v_{if} + P_{2i}v_{iu}][1 + P_{3i}k_i/(k_i - 1)]$.

Eq. (2) symbolizes the likelihood of cluster i being encompassed within the sample.

$$\alpha_k = 1 - \left(\binom{N - W_k}{n_1} \middle/ \binom{N}{n_1}\right) \qquad (2)$$

The probability of cluster i being included in the sample is defined and specified using (SRS) without replacement.:

$$\alpha_k = 1 - \left(1 - \frac{W_k}{N}\right)^{n_1}$$

where, W_k denote the numbers of unit k on the network.

Variance of an unbiased estimator:

$$Var_1\left(\frac{1}{N}\sum_{k=1}^{\tau}\frac{\hat{v}_{kf}j_k}{\alpha_k}\right) = \frac{1}{N^2}\sum_{j-1}^{\xi}\sum_{h=1}^{\xi}v_{jf*}v_{hf*}(v_{kb} - v_kv_b)/(v_kv_kv_{kb}) \qquad (3)$$

with $E_1\left[\frac{1}{N^2}\sum_{k=1}^{\tau}m_k\frac{1}{\alpha_k^2}\left(\frac{(P_{1k}v_{kf}+P_{2k}v_{ku})}{P_{1k}^2(1+P_{3k}k_k/(k_k-1))}\right)\right]$,
and

$$E_1Var_2(\hat{v}_{fHT}) = \frac{1}{N^2}\sum_{k=1}^{N}m_k\frac{\psi_k}{\alpha_k}.$$

where,

$$\psi_k = \left[\frac{P_{1k}v_{kf}+P_{2k}v_{ku}}{P_{1k}^2(1+P_{3k}k_k/(k_k-1))}\right] \qquad (4)$$

The variance from unknown is as:

$$V(\hat{v}_{fHT}) = \frac{1}{N^2}\left[\sum_{k=1}^{K}\sum_{b=1}^{K}v_{kf^*}v_{bf^*}(v_{kb} - v_kv_b)\middle/(v_kv_bv_{kb}) + \sum_{k=1}^{N}m_k\frac{\phi_k}{\alpha_k}\right] \qquad (5)$$

with, $\phi_k = (P_{1k}H_kT_{2k} + T_{1k}Q_kP_{2k})v_f - \binom{P_{2k}H_kT_{2k}}{+T_{2k}Q_kP_{2k}}v_b - 2T_{2k}P_{2k}(T_{1k}P_{1k}v_{kf} + T_{2k}P_{2k}v_{ku}), H_k = \frac{T_{2k}}{1+P_{3k}k_k/(k_k-1)}$,
and

$$Q_k = \frac{P_{2k}}{T_{1k}^2(1+P_{3k}k_k/(k_k-1))}.$$

2. THE PROPOSED MODEL

The new model is a hybrid of randomised response model and adaptive two-stage cluster sampling, to estimate rare sensitive attributes. We provide estimation method for the mean of the population with a rare sensitive attribute (MNSA) using modified Horvitz-Thompson type estimator. We investigate the known and unknown conditions of a unique non-sensitive trait that is unrelated. In the second stage of sampling, we collect responses from the elementary units using the randomization method suggested by Singh and Tarray [24], as well as the recent work [25, 26].

2.1 When the rare attribute that is unrelated is known

Participants are requested to engage with and encounter the randomization device without prior knowledge of whether the distinctive characteristic will be revealed or concealed, contingent upon the identification of the proportion of individuals possessing the unrelated rare attribute.

The respondents selected from the ith cluster would be given a deck of k_i cards as the randomization mechanism and card picked has one of the questions to determine if the respondent has (i) rare sensitive attribute be "F" with probability P_{1i}, (ii) rare unrelated attribute "U" with probability P_{2i}, and (iii) Blank Card.

The respondent must repeat the preceding steps without taking the card out if the statement (iii) is drawn. When the statement (iii) is drawn, he or she is compelled to report "No" at the second level. The suggested deck's total number of cards and the likelihood that "Yes" will be answered are both provided.

$$\phi_{i0} = [P_{1i}\beta_{if} + P_{2i}\beta_{iu}] + P_{2i}(1-\beta_{if}) \qquad (6)$$

Let's assume that, $m_i \to \infty, m_i\phi_{i0} = v_{i0} > 0$, as $\phi_{i0} \to 0, m_i\beta_{if} = v_{if} > 0, as \beta_{if} \to 0$, and $m_i\beta_{iu} = v_{iu} > 0, as \to \infty, \beta_{iu} \to 0$, where,

$$v_{i0} = [P_{1i}v_{if} + P_{2i}v_{iu}] + P_{2i}(1-v_{if}) \qquad (7)$$

$$f(v_{i0}) = \prod_{j=1}^{m_i} \frac{e^{-v_{i0}}v_{i0}{}^{y_{ij}}}{y_{ij}!}, \qquad (8)$$

Let $Y \sim Pois\left(v_{i0}\right)$ a random sample of m_i from the ith cluster. $L(v_{i0}) = e^{-m_i v_{i0}}v_{i0}{}^{\sum_{j=1}^{m_i} y_{ij}} \prod_{j=1}^{m_i} \frac{1}{y_{ij}!}$,

where, Eq. (8) is the likelihood function.
From Eq. (7) and Eq. (8), the MLE v_{if}

$$\hat{v}_{if} = \frac{1}{P_{1i}}\left[\frac{\sum_{j=1}^{m_i} y_{ij}}{P_{2i}(1-v_{iu})} - P_{2i}v_{iu}\right] \qquad (9)$$

To show that $\hat{v}_{if}$ is unbiased parameter v_{if}, following Eq. (2) check [7, 25].

Theorem 1: The estimator $\hat{v}_{if}$ is an unbiased estimator of the parameter v_{if}.

Proof: following Eq. (2) [24].

If for any cluster, $w_k \neq 1$, and indicator variable (IV)=0, it implies that the kth cluster was not chosen in the initial because it falls short of the requirement. So, $j_k=1$. A the two-stage adaptive sampling strategy, a improved type of Horvitz-Thompson estimator of MNSA in the population is represented as:

$$\hat{v}_{HT} = \frac{1}{N}\sum_{k=1}^{\tau} \frac{\hat{v}_{kf}j_k}{\alpha_k} \qquad (10)$$

The estimator $\hat{v}_{HT}$ is computed as a weighted sum of the observed values, where the weights are inversely proportional to the inclusion probabilities, ensuring unbiased estimation of the population total.

Theorem 2: The estimator $\hat{v}_{HT}$ of the MNSA v_f is unbiased.

Proof: consider

$$E_1E_2(\hat{v}_{HT}) = E_1E_2\left(\frac{1}{N}\sum_{k=1}^{\tau} \frac{\hat{v}_{kf}j_k}{\alpha_k}\right)$$

where, E_1 is the anticipated total number of first-round picks and E_2 is the anticipated total number of second-round picks.

$$E_1E_2(\hat{v}_{HT}) = E_1\left(\frac{1}{N}\sum_{k=1}^{\tau} \frac{\hat{v}_{kf}j_k}{\alpha_k}\right) \qquad (11)$$

See Mansour (2021).

$$E_1E_2(\hat{v}_{HT}) = \left(\frac{1}{N}\sum_{k=1}^{\tau} v_{if}\right) = v_f \qquad (12)$$

Theorem 3: The variance of the estimator $\hat{v}_{fHT}$ is:

$$V(\hat{v}_{HT}) = \frac{1}{N^2}\left[\sum_{k=1}^{K}\sum_{b=1}^{K} v_{kf*}v_{bf*}(v_{kb}-v_kv_b)/(v_kv_bv_{kb}) + \sum_{k=1}^{N} m_k \frac{\psi_k}{\alpha_k}\right],$$

where,

$$\psi_{rk} = \left[\frac{P_{1k}v_{kf}+P_{2k}v_{ku}+P_{2k}(1-v_{ku})}{P_{1k}^2 v_{kf}-P_{2k}^2 v_{ku}}\right] \qquad (13)$$

Proof. The detailed proof can be found in the study by Singh and Tarray [24]. The abridged proof is as follows:

$$Var(\hat{v}_{HT}) = Var_1E_2(\hat{v}_{HT}) + E_1Var_2(\hat{v}_{HT}) \qquad (14)$$

where, var_1 − variance of the adaptive, and var_2 − variance over the second stage
The first term of (14) is:

$$Var_1E_2(\hat{v}_{HT}) = Var_1E_2\left(\frac{1}{N}\sum_{k=1}^{\tau} \frac{\hat{v}_{kf}j_k}{\alpha_k}\right) = Var_1\left(\frac{1}{N}\sum_{k=1}^{\tau} \frac{\hat{v}_{kf}j_k}{\alpha_k}\right) \qquad (15)$$

Thus,

$$Var_1\left(\frac{1}{N}\sum_{k=1}^{\tau} \frac{\hat{v}_{kf}j_k}{\alpha_k}\right) = \frac{1}{N^2}\sum_{j-1}^{\xi}\sum_{h=1}^{\xi} v_{jf*}v_{hf*}cov(v_j,v_h)/(v_jv_h) \\ Var_1\left(\frac{1}{N}\sum_{k=1}^{\tau} \frac{\hat{v}_{kf}j_k}{\alpha_k}\right) = \frac{1}{N^2}\sum_{j-1}^{\xi}\sum_{h=1}^{\xi} v_{jf*}v_{hf*}(v_j,v_h)/(v_jv_h) \qquad (16)$$

Variance of an unbiased estimator:

$$Var_1\left(\frac{1}{N}\sum_{k=1}^{\tau} \frac{\hat{v}_{kf}j_k}{\alpha_k}\right) = \frac{1}{N^2}\sum_{j-1}^{\xi}\sum_{h=1}^{\xi} v_{jf*}v_{hf*}(v_{kb}-v_kv_b)/(v_kv_kv_{kb}) \qquad (17)$$

The second term of (14) is:

$$E_1Var_2(\hat{v}_{fHT}) = E_1Var_2\left(\frac{1}{N}\sum_{k=1}^{\tau} \frac{\hat{v}_{kf}j_k}{\alpha_k}\right)$$
$$= E_1\left(\frac{1}{N^2}\sum_{k=1}^{\tau} \frac{j_kVar_2(\hat{v}_{kf})}{\alpha_k^2}\right)$$
$$\hat{v}_{if} = \frac{1}{P_{1i}}\left[\frac{\sum_{j=1}^{m_i} y_{ij}}{P_{2i}(1-v_{iu})} - P_{2i}v_{iu}\right]$$
$$= E_1\left[\frac{1}{N^2}\sum_{k=1}^{\tau} \frac{j_k}{\alpha_k^2}Var_2\left(\frac{1}{P_{1k}}\left(\frac{\sum_{j=1}^{m_i} y_{ij}}{m_kP_{2i}(1-v_{iu})} - P_{2i}v_{iu}\right)\right)\right]$$
$$= E_1\left[\frac{1}{N^2}\sum_{k=1}^{\tau} \frac{j_k}{\alpha_k^2}\left(\frac{1}{P_{1k}^2}\left(\frac{\sum_{j=1}^{m_k} Var_2(y_{kj})}{m_k^2P_{2i}(1-v_{iu})}\right)\right)\right]$$
$$= E_1\left[\frac{1}{N^2}\sum_{k=1}^{\tau} \frac{j_k}{\alpha_k^2}\left(\frac{1}{P_{1k}^2}\left(\frac{\sum_{j=1}^{m_k} v_{k0}}{m_k^2P_{2i}(1-v_{iu})^2}\right)\right)\right]$$
$$= E_1\left[\frac{1}{N^2}\sum_{k=1}^{\tau} \frac{j_k}{\alpha_k^2}\left(\frac{1}{P_{1k}^2}\left(\frac{v_{k0}}{m_kP_{2i}(1-v_{iu})^2}\right)\right)\right] \qquad (18)$$

$$= E_1\left[\frac{1}{N^2}\sum_{k=1}^{\tau}\frac{j_k}{\alpha_k^2}\left(\frac{1}{P_{1k}^2}\left(\frac{(P_{1k}v_{kf}+P_{2k}v_{ku})P_{2i}(1-v_{iu})}{m_kP_{2i}(1-v_{iu})^2}\right)\right)\right]$$

$$= E_1\left[\frac{1}{N^2}\sum_{k=1}^{\tau}m_k\frac{1}{\alpha_k^2}\left(\frac{(P_{1k}v_{kf}+P_{2k}v_{ku})}{P_{1k}^2P_{2i}(1-v_{iu})}\right)\right]$$

$$E_1Var_2(\hat{v}_{fHT}) = \frac{1}{N^2}\sum_{k=1}^{N}m_k\frac{\psi_k}{\alpha_k} \qquad (19)$$

where,

$$\psi_{rk} = \left[\frac{P_{1k}v_{kf}+P_{2k}v_{ku}}{P_{1k}^2P_{2i}(1-v_{iu})}\right].$$

The $\hat{v}_{fHT}$ in Eq. (13) earlier given was obtained by substituting the equation in (11) and (19) into (14).

2.2 The case when unrelated rare non-sensitive attribute is unknown

In the case of randomized response (RR) strategies combined with adaptive cluster sampling (ACS), the use of an unrelated rare non-sensitive attribute can play a crucial role in maintaining respondent confidentiality and enhancing the accuracy of data collection. Typically, RR techniques rely on respondents answering questions about sensitive attributes indirectly, often by referencing unrelated non-sensitive attributes to introduce randomness.

Unknown rare non-sensitive attributes can lead to increased response bias, difficulty in designing the RR mechanism, complexity in data analysis, and increased operational challenges. Respondents may feel less secure about the confidentiality of their responses, potentially undermining the effectiveness of the RR strategy. Designing the RR mechanism requires precise probability distributions, which can be challenging when the attribute is unknown. Data analysis becomes more complex, potentially leading to biased or inaccurate conclusions. To address these challenges, innovative approaches include pre-survey piloting, statistical adjustment techniques, and adaptive RR mechanisms.

Each respondent is asked twice for a response in order to estimate the MNSA. While the unrelated, unusual, harmless trait is unidentified. These randomization tools are made up of decks of cards and cards that are comparable, as explained in Section 2.1. The initial randomization device based on the ith cluster's respondents asks them to respond "yes" or "no" at start. We follow the procedure outlined in section 2.1.

Using a second randomization system consisting of the outlined statements with probabilities T_{1i}, T_{2i}, and T_{3i} instead of P_{1i}, P_{2i}, and P_{3i}, the respondents chosen from the ith cluster are required to answer the same questions. Responders in the ith cluster have "yes" responses with the following probabilities:

$$\phi_{i1} = \left[P_{1i}\beta_{if} + P_{2i}\beta_{iu}\right] + P_{2i}(1-\beta_{iu}) \qquad (20)$$

and

$$\phi_{i2} = \left[T_{1i}\beta_{if} + T_{2i}\beta_{iu}\right] + T_{2i}(1-\beta_{iu}) \qquad (21)$$

If $m_i \to \infty$, $\phi_{i1} \to \infty$, and $\phi_{i2} \to 0$, $m_i\phi_{i1} = v_{i1} > 0$, and $m_i\phi_{i2} = v_{i2} > 0$, as $\beta_{iu} \to 0$ where $m_i\beta_{iu} = v_{iu} > 0$, and $m_i\beta_{if} = v_{if} > 0$. Subsequently, (20) and (21) $m_i\beta_{iu} = v_{iu} > 0$.

where,

$$\upsilon_{i1} = \left[P_{1i}\upsilon_{if} + P_{2i}\upsilon_{iu}\right] + P_{2i}(1-\beta_{iu}) \qquad (22)$$

and

$$\upsilon_{i2} = \left[T_{1i}\upsilon_{if} + T_{2i}\upsilon_{iu}\right] + T_{2i}(1-\beta_{iu}) \qquad (23)$$

Simplifying (16) and (17) results in

$$\frac{1}{m_i}\sum_{j=1}^{m_i}y_{i1j} = \left[P_{1i}\hat{v}_{if} + P_{2i}\hat{v}_{iu}\right] + P_{2i}(1-\beta_{iu}) \qquad (24)$$

$$\frac{1}{m_i}\sum_{j=1}^{m_i}y_{i2j} = \left[T_{1i}\hat{v}_{if} + T_{2i}\hat{v}_{iu}\right] + T_{2i}(1-\beta_{iu}) \qquad (25)$$

Solving Eq (24) and Eq (25), we then have the estimators of v_{if} and v_{iu} to be:

$$\hat{v}_{if2} = \frac{1}{m_i(T_{2i}P_{1i}+T_{1i}P_{2i})}\left(\frac{T_{2i}\sum_{j=1}^{m_i}y_{i1j}}{P_{2i}(1-\beta_{iu})} - \frac{P_{2i}\sum_{j=1}^{m_i}y_{i2j}}{T_{2i}(1-\beta_{iu})}\right) \qquad (26)$$

where, $T_{2i}P_{1i} \neq T_{1i}P_{2i}$.

$$\hat{v}_{iu2} = \frac{1}{m_i(T_{1i}P_{2i}+T_{2i}P_{1i})}\left(\frac{T_{1i}\sum_{j=1}^{m_i}y_{i1j}}{P_{2i}(1-\beta_{iu})} - \frac{P_{1i}\sum_{j=1}^{m_i}y_{i2j}}{T_{2i}(1-\beta_{iu})}\right) \qquad (27)$$

where, $T_{2i}P_{1i} \neq T_{1i}P_{2i}$.

The computation of the estimator for the population mean of the overall count of individuals with a rare and sensitive trait is subsequently carried out using the designated method.

$$\hat{v}_{HT2} = \left(\frac{1}{N}\sum_{k=1}^{\tau}\frac{\hat{v}_{kf2}j_k}{\alpha_k}\right) \qquad (28)$$

Theorem 4: The estimator or $\hat{v}_{fHT2}$ of the MNSA attribute is unbiased.
Proof.
Since $E(E_1) = E(E_2)$ we conclude that $E(\hat{v}_{fHT2}) = v_f$. where, E_1, and E_2 are the expected total number of first and second -stage selections receptively and $E(\cdot)$ is the expectation.

The variance of the estimator $\hat{v}_{fHT2}$ is:

$$V(\hat{v}_{HT2}) =$$
$$\frac{1}{N^2}\left[\sum_{k=1}^{K}\sum_{b=1}^{K}\frac{v_{kf^*}v_{bf^*}(v_{kb}-v_kv_b)/(v_kv_bv_{kb})}{+\sum_{k=1}^{N}m_k\frac{\phi_k}{\alpha_k}}\right], \qquad (29)$$

This equation represents the variance $V(\hat{v}_{HT2})$ of the Horvitz-Thompson estimator (HT2) for a Poisson-distributed adaptive cluster sampling model. The first term inside the brackets involves double summation over clusters, accounting for interactions between pairs of clusters k and b. The second term sums over all clusters, adjusting for specific sampling parameters m_k, ϕ_k and α_k.

$$H_k = \frac{T_{2k}}{P_{2i}(1-\beta_{iu})},$$

$$Q_k = \frac{P_{2k}}{T_{2i}(1-\beta_{iu})}$$

These equations define two parameters, H_k and Q_k, used in the adaptive cluster sampling model. Here, T_{2k} and P_{2k} rrepresents specific values related to cluster k, while P_{2i} and T_{2i} are reference values for another index i. The term $(1 - \beta_{iu})$ adjusts for a factor β_{iu} influencing the relationship between clusters and the overall sample.

the estimators specified in Eqs (4) and (13). These variances provide insight into the precision and reliability of the population mean estimates derived from the sampling models. For clarity and ease of comparison, the results of these calculations have been organized and presented in Table 1. The tabulated variances serve as a critical component in assessing the effectiveness of the proposed sampling models and their respective estimators in capturing the true population parameters.

3. APPLICATIONS

In this section the existing model and the proposed models were used for estimation and compared in Table 1 and Table 2 respectively. The comparison consists of six clusters. From each cluster, a sizable random sample was collected, and the MNSA was calculated. The estimated - parameters of which are {1,0,1,0,3,5}. Each unit (whether there are one or two) includes both adjacent units. $n_1=2$ is the starting sample size. There are $\binom{6}{2} = 15$ potential samples with a chance of 1/15 using the adaptive design, where the initial sample is chosen by SRS without replacement. Considering the population from [19], Table 1 shows the produced observations together with the values of every estimator. The condition is indicated by $v_{kf^*} < 1, v_b + v_k = 1, v_{bf^*} + v_{kb} = 1$.

From the above Table 1, ψ_k and ψ_{rk} were obtained from the Eqs (4) and (13).

The variances of population mean were obtained based on the estimators stated Eqs. (5) and (14) is tabulated in Table 2 as follows:

From the above Table 2 for different values v_{kf^*}, v_{bf^*}, v_{kb}, v_k and v_b we have obtained the values of ψ_k and ψ_{rk}. From Eqs. (5) and (14) were obtained variances of existing model and proposed model, the variance $V(\hat{v}_{HT})$ under proposed model is less than the variance $V(\hat{v}_{fHT})$ of existing model. Which results gain in efficiency. Hence our proposed model works well as compared to existing literature.

3.1 Empirical study

The variances of the population mean were calculated using

Table 1. Observation result and the estimator values

Networks	v_{kf^*}	v_{bf^*}	v_{kb}	v_k	v_b	ψ_k	ψ_{rk}
1,0	0.3	0.7	0.3	0.3	0.7	0.888	0.81
1,1	0.4	0.8	0.2	0.2	0.8	1.008	1.26
1,0	0.5	0.9	0.1	0.1	0.9	1.128	1.71
1,3;0.5	0.4	0.6	0.4	0.4	0.6	0.768	1.08
1,5;3,0	0.4	0.4	0.5	0.5	0.5	0.528	0.99
0,1	0.8	0.8	0.1	0.1	0.9	1.008	2.79
0,0	0.9	0.9	0.1	0.9	0.8	1.128	3.15
0,3;0.5	0.7	0.7	0.2	0.2	0.8	0.888	2.34
0,5;3	0.5	0.5	0.4	0.4	0.6	0.648	1.44
1,0	0.6	0.6	0.3	0.3	0.7	0.768	1.89
1,3;0,5	0.5	0.5	0.5	0.5	0.5	0.648	1.35
1,5;0,3	0.6	0.4	0.6	0.6	0.4	0.528	1.62
0,3;5	0.7	0.3	0.7	0.7	0.3	0.408	1.89
0,5;3	0.9	0.3	0.6	0.6	0.4	0.408	2.7
3,5;0	0.8	0.2	0.7	0.7	0.3	0.288	2.25
Mean	0.631579	0.478947	0.478947	0.521053	0.515789	0.631579	0.478947

Table 2. The variance of existing and the proposed estimator and relative precision

v_{kf^*}	v_{bf^*}	v_{kb}	v_k	v_b	ψ_k	ψ_{rk}	$V(\hat{v}_{fHT})$ Existing	$V(\hat{v}_{HT})$ Proposed	Efficiency
0.3	0.7	0.3	0.3	0.7	0.888	0.81	267.18	227.82	117.27
0.4	0.8	0.2	0.2	0.8	1.008	1.26	261.52	188.29	138.89
0.5	0.9	0.1	0.1	0.9	1.128	1.71	257.07	169.56	151.60
0.4	0.6	0.4	0.4	0.6	0.768	1.08	411.9	219.68	187.50
0.4	0.4	0.5	0.5	0.5	0.528	0.99	649.09	239.65	270.84
0.8	0.8	0.1	0.1	0.9	1.008	2.79	444.58	160.60	276.82
0.9	0.9	0.1	0.9	0.8	1.128	3.15	444.06	159.01	279.25
0.7	0.7	0.2	0.2	0.8	0.888	2.34	474.99	168.96	281.11
0.5	0.5	0.4	0.4	0.6	0.648	1.44	569.56	201.36	282.84
0.6	0.6	0.3	0.3	0.7	0.768	1.89	514.89	181.31	283.97
0.5	0.5	0.5	0.5	0.5	0.648	1.35	610.25	214.7	284.10
0.6	0.4	0.6	0.6	0.4	0.528	1.62	898.74	211.54	424.85
0.7	0.3	0.7	0.7	0.3	0.408	1.89	1356.9	209.22	648.54
0.9	0.3	0.6	0.6	0.4	0.408	2.7	1550.78	185.50	835.99
0.8	0.2	0.7	0.7	0.3	0.288	2.25	2105.43	199.19	1056.95
0.8	0.2	0.8	0.8	0.2	0.288	2.16	2197.00	207.52	1058.65
0.6	0.1	0.9	0.9	0.1	0.168	1.35	3295.66	254.08	1297.07
0.7	0.1	0.8	0.8	0.2	0.168	1.8	3452.47	219.79	1570.80
0.9	0.1	0.9	0.9	0.1	0.168	2.43	4237.33	206.45	2052.42

4. CONCLUSIONS

This research endeavors to bridge the gap between adaptive sampling techniques and respondent confidentiality, presenting a novel Poisson-distributed adaptive cluster sampling model enhanced by a randomized response strategy. Through theoretical development and empirical validation, we aim to demonstrate the efficacy of this integrated approach in yielding high-quality, reliable data from clustered and sensitive populations.

In conclusion, the utilization of cluster sampling and the randomized response technique, coupled with the introduction of the adaptive cluster sampling randomized response model featuring a Poisson distribution, has brought forth valuable insights into estimating the prevalence of sensitive groups within the population. The demonstrated lower variance in the proposed model not only signifies increased efficiency but also highlights its potential for more accurate estimations. this study employed cluster sampling and the randomized response technique to gauge the proportion of the population associated with a sensitive group. Introducing the adaptive cluster sampling randomized response model with a Poisson distribution proved to be a significant advancement. The results showcased that this proposed model exhibited lower variance, indicating greater efficiency compared to existing models. Notably, the study demonstrated the superiority of the proposed model over its counterparts, marking a substantial contribution to the field. This innovative methodology holds promise for advancing the field of statistical sampling and improving the outcomes of various research endeavors reliant on accurate and unbiased data.

REFERENCES

[1] Warner, S.L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. Journal of the American Statistical Association, 60(309): 63-69. https://doi.org/10.1080/01621459.1965.10480775

[2] Greenberg, B.G., Abul-Ela, A.L.A., Simmons, W.R., Horvitz, D.G. (1969). The unrelated question randomized response model: Theoretical framework. Journal of the American Statistical Association, 64(326): 520-539. https://doi.org/10.1080/01621459.1969.10500991

[3] Chaudhuri, A., Mukerjee, R. (1987). Randomized response: Theory and techniques. Routledge. https://doi.org/10.1201/9780203741290

[4] Blair, G., Imai, K., Zhou, Y.Y. (2015). Design and analysis of the randomized response technique. Journal of the American Statistical Association, 110(511): 1304-1319. https://doi.org/10.1080/01621459.2015.1050028

[5] Tracy, D.S., Mangat, N.S. (1996). Some development in randomized response sampling during the last decade-a follow up of review by Chaudhuri and Mukerjee. Journal of Applied Statistical Science, 4(2/3): 147-158.

[6] Singh, S. (2003). Advanced sampling theory with applications. Dordrecht. Netherlands: Kiuwer Academic Publishers. https://doi.org/10.1007/978-94-007-0789-4

[7] Dare, J.R., Agunbiade, D.A., Famurewa, O.K., Adesina, O.S., Adedotun, D.F., Iyaniwura, O. (2018). Approximation techniques for maximizing likelihood functions of generalized linear mixed models for binary response data. International Journal of Engineering & Technology, 7(4): 4911-4917. 10.14419/ijet.v7i4.24842

[8] Adedotun, A.F., Odusanya, O.A., Okagbue, H.I., Ogundile, O.O. (2022). Analysis of reported cases of diabetes disease in Nigeria: A Survival Analysis Approach. Planning, 17(2): 643-647. https://doi.org/10.18280/ijsdp.170229

[9] Hong, K.H., Yum, J.K., Lee, H.Y. (1994). A stratified randomized response technique. Korean Journal of Applied Statistics, 7(1): 141-147.

[10] Kim, J.M., Warde, W.D. (2005). Some new results on the multinomial randomized response model. Communications in Statistics-Theory and Methods, 34(4): 847-856. https://doi.org/10.1081/STA-200054378

[11] Kim, J.M., Warde, W.D. (2004). A stratified Warner's randomized response model. Journal of Statistical Planning and Inference, 120(1-2): 155-165. https://doi.org/10.1016/S0378-3758(02)00500-1

[12] Mangat, N.S., Singh, R. (1990). An alternative randomized response procedure. Biometrika, 77(2): 439-442. https://doi.org/10.1093/biomet/77.2.439.

[13] Mangat, N.S. (1994). An improved randomized response strategy. Journal of the Royal Statistical Society: Series B (Methodological), 56(1): 93-95. https://doi.org/10.1111/j.2517-6161.1994.tb01962.x

[14] Kim, J.M., Tebbs, J.M., An, S.W. (2006). Extensions of Mangat's randomized-response model. Journal of Statistical Planning and Inference, 136(4): 1554-1567. https://doi.org/10.1016/j.jspi.2004.10.005

[15] Kim, J.M., Warde, W.D. (2005). A mixed randomized response model. Journal of Statistical Planning and Inference, 133(1): 211-221. https://doi.org/10.1016/j.jspi.2004.03.011

[16] Lee, G.S., Uhm, D., Kim, J.M. (2013). Estimation of a rare sensitive attribute in a stratified sample using Poisson distribution. Statistics, 47(3): 575-589. https://doi.org/10.1080/02331888.2011.625503

[17] Land, M., Singh, S., Sedory, S.A. (2012). Estimation of a rare sensitive attribute using Poisson distribution. Statistics, 46(3): 351-360. https://doi.org/10.1080/02331888.2010.524300

[18] Suman, S., Singh, G.N. (2019). An ameliorated stratified two-stage randomized response model for estimating the rare sensitive parameter under Poisson distribution. Statistics, 53(2): 395-416. https://doi.org/10.1080/02331888.2019.1569665

[19] Singh, H.P., Tarray, T.A. (2016). A stratified Tracy and Osahan's two-stage randomized response model. Communications in Statistics-Theory and Methods, 45(11): 3126-3137. https://doi.org/10.1080/03610926.2014.895839

[20] Akingbade, T.J., Okafor, F.C. (2019). Generalized class of difference-type ratio estimators for estimating the population mean using known population parameter of auxiliary variable. Pakistan Journal of Statistics, 35(3): 197-215.

[21] Akingbade, T.J., Okafor, F.C. (2019). A class of ratio-type estimator using two auxiliary variables for

estimating the population mean with some known population parameters. Pakistan Journal of Statistics and Operation Research, 329-340. https://doi.org/10.18187/pjsor.v15i2.2558

[22] Ajayi, A.O., Ayeleso, T.O., Gboyega, A.F., Adesina, O.S. (2023). A modified generalized class of exponential ratio type estimators in ranked set sampling. Scientific African, 19: e01447. https://doi.org/10.1016/j.sciaf.2022.e01447

[23] Thompson, S.K. (2017). Adaptive and network sampling for inference and interventions in changing populations. Journal of Survey Statistics and Methodology, 5(1): 1-21. https://doi.org/10.1093/jssam/smw035

[24] Singh, H.P., Tarray, T.A. (2014). A dexterous randomized response model for estimating a rare sensitive attribute using Poisson distribution. Statistics & Probability Letters, 90: 42-45. https://doi.org/10.1016/j.spl.2014.03.019

[25] Mansour, M.M., Enayat, M. (2021). Adaptive cluster sampling randomized response model with electronically application. Journal of Nonlinear Sciences and Applications, 14: 8-14. http://dx.doi.org/10.22436/jnsa.014.01.02

[26] Lee, G.S., Son, C.K. (2022). An estimation of a sensitive attribute using adjusted Kuk's randomization device with stratified unequal probability sampling. Communications in Statistics-Theory and Methods, 51(1): 1-25. https://doi.org/10.1080/03610926.2020.1821890