

Article

Not peer-reviewed version

---

# A Refined Regression Estimator for General Inverse Adaptive Cluster Sampling

---

[Nipaporn Chutiman](#) , [Supawadee Wichitchan](#) , [Chawalit Boonpok](#) , [Monchaya Chiangpradit](#) , [Pannarat Guayjarempanishk](#) \*

Posted Date: 30 October 2025

doi: 10.20944/preprints202510.2340.v1

Keywords: adaptive cluster sampling; general inverse sampling; regression estimator; auxiliary information



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# A Refined Regression Estimator for General Inverse Adaptive Cluster Sampling

Nipaporn Chutiman <sup>1</sup>, Supawadee Wichitchan <sup>1</sup>, Chawalit Boonpok <sup>1</sup>, Monchaya Chiangpradit <sup>1</sup> and Pannarat Guayjarernpanishk <sup>2,\*</sup>

<sup>1</sup> Department of Mathematics, Faculty of Science, Mahasarakham University, Maha Sarakham, 44150, Thailand

<sup>2</sup> Department of Mathematics, Faculty of Science, Khon Kaen University, Khon Kaen, 40002, Thailand

\* Correspondence: panngu@kku.ac.th

## Abstract

Adaptive cluster sampling (ACS) is a sampling method commonly employed when the population is rare and exhibits clustering. However, the initial sample selection may include units that do not satisfy the specified condition. To address this, general inverse sampling is incorporated into ACS, where the initial units are selected sequentially and termination criteria are applied to regulate the number of rare elements drawn from the population. The objective of this study is to develop an estimator of the population mean by utilizing auxiliary information within the framework of general inverse adaptive cluster sampling. The proposed estimator, constructed on the basis of a regression-type estimator, is analytically examined. A simulation study was conducted to validate the theoretical results. In this study, the region of interest was divided into 400 square units (20 rows by 20 columns). The results demonstrate that the proposed estimator, which incorporates auxiliary variables, consistently yields a lower variance than the conventional mean estimator without auxiliary information. This superiority holds across all scenarios considered, specifically when the predetermined number of rare units  $r$  ranges from two to ten. Therefore, the proposed estimator is shown to be more efficient than the estimator that does not employ auxiliary information.

**Keywords:** adaptive cluster sampling; general inverse sampling; regression estimator; auxiliary information

**MSC:** 62D05; 62F40

## 1. Introduction

Adaptive cluster sampling (ACS) has been recognized as an effective approach for studying population that is both rare and spatially clustered, with its introduction attributed to Thompson [1]. Initially, a set of initial sample units is drawn by employing simple random sampling without replacement (SRSWOR). Whenever one of these units meets the predefined criterion  $C$  for the variable of interest, the neighboring units of that sample are included. The procedure is then extended iteratively, such that any further units fulfilling the condition will result in the inclusion of their neighborhoods as well. This procedure is repeated until no further units satisfy the specified condition. In addition, if the initial unit does not satisfy condition  $C$  for the variable of interest, no augmentation occurs and the cluster size is one. The set of initial sample units selected and all neighborhood units that satisfy the condition is referred to as a network. ACS has been extensively applied in situations where populations are rare and exhibit clustered distributions. Its applications include forest ecosystem monitoring [2], herpetofauna surveys in tropical rainforests [3], assessments of sea lamprey larval distributions [4], and investigating freshwater mussel populations [5]. The method has also been used in hydroacoustic investigations [6] and research related to the COVID-19

pandemic [7-9]. Beyond ecological and epidemiological contexts, ACS has been further adapted to emerging domains, such as autonomous technologies and IoT-driven applications [10-11].

In many studies, the efficiency of estimation under ACS has been improved through the incorporation of auxiliary variables correlated with the variable of interest. A ratio-type estimator was introduced in [12], and subsequent modifications were proposed in [13]. An extension incorporating two auxiliary variables was presented in [14]. Further developments included ratio estimators based on known population parameters of the auxiliary variable, such as the coefficient of variation, kurtosis, and skewness [15,16]. Generalized exponential-type estimators were proposed in [17,18], while a generalized class of ratio-type estimators was developed in [19].

However, in ACS, the initial sample size is determined before sample selection, and it may happen that not all randomly selected units meet the specified condition  $C$ . Inverse ACS was proposed in [20], in which inverse sampling is applied within the ACS framework. In this design, the initial units are selected sequentially, and the sampling process is terminated according to stopping rules established to regulate the number of rare elements obtained from the population. However, in some cases of inverse sampling, the procedure may not be completed if the number of units meeting the specified condition is very small, while the stopping rules are set higher than the number of qualifying units. To address this limitation, general inverse sampling was introduced in [21], providing a more suitable approach for practical applications and subsequently applied to ACS. An estimator in general inverse ACS based on the Rao–Blackwell theorem was developed in [22], and the integration of unequal probability inverse sampling with ACS was investigated in [23].

As mentioned above, general inverse ACS has thus far been studied only in relation to the variable of interest. In certain cases, however, other variables correlated with the variable of interest may also be available, and their data can often be collected simultaneously without additional cost. Incorporating auxiliary variable information is a common strategy to enhance estimation precision. An estimator utilizing auxiliary variable information, specifically a regression-type estimator in general inverse sampling, was proposed in [24]. In this paper, an estimator was developed using auxiliary information in general ACS. Section 2 presents inverse sampling and general inverse sampling. Section 3 describes general inverse adaptive cluster sampling. Section 4 introduces the proposed estimator that uses auxiliary variable information in general inverse adaptive cluster sampling. Simulation studies are presented in Section 5, and finally, Section 6 presents the conclusions of the study.

## 2. Inverse Sampling and General Inverse Sampling

Population size  $N$  units consist of the set of values  $U = (y_1, y_2, \dots, y_N)$ . Let  $Y$  be the variable of interest and  $y_i$  is the  $y$ -value associated with the  $i^{\text{th}}$  unit. The population is divided into two subgroups according to whether the  $y$ -value satisfies the pre-condition  $C$ . Following the notation of [20], define the two subgroups by  $U_M = \{y : y_i \in C ; i = 1, 2, \dots, N\}$  and  $U_{N-M} = U_{M'} = \{y : y_i \notin C ; i = 1, 2, \dots, N\}$ , where  $M$  is the unknown number of units of  $U_M$ . In this sampling framework, the classification of a unit into subgroups is not determined in advance but becomes known only upon its selection. The sampling procedure involves sequentially selecting units at random from the target population. Sampling continues until a predetermined number of  $r$  units ( $1 < r \leq M$ ) from  $U_M$  are sampled.

The estimator in inverse sampling was presented in [20]. Suppose a sample size  $n_I$  is selected by SRSWOR. If at least  $r$  ( $r > 1$ ) units of  $U_M$  are observed, then no further sampling is conducted. Otherwise, sampling continues sequentially until exactly  $r$  units are selected. The total sequential sample size is defined as  $n_T$ . Then an unbiased estimator of  $\mu_y$  is:

$$\bar{y}_{IS} = \begin{cases} \frac{1}{n_T} \sum_{i=1}^{n_T} y_i & \text{if } n_T = n_I, \\ \frac{1}{N} (M\bar{y}_M + (N-M)\bar{y}_{M'}) & \text{if } n_T > n_I, \end{cases} \quad (1)$$

where  $\bar{y}_M = \frac{1}{r} \sum_{i \in S_M} y_i$ ;  $S_M$  is the index label of the members of  $U_M$ .

$\bar{y}_{M'} = \frac{1}{n_T - r} \sum_{i \in S_{M'}} y_i$ ;  $S_{M'}$  is the index label of the members of  $U_{M'}$ . In application,  $M$  will be

not known. The unbiased estimator of  $M$  is  $\hat{M} = \frac{N(r-1)}{(n_T-1)}$ .

The variance of  $\bar{y}_{IS}$  is:

$$V(\bar{y}_{IS}) = \begin{cases} \frac{(N-n_I)S^2}{N n_I} & \text{if } n_T = n_I, \\ V(\bar{y}_I) & \text{if } n_T > n_I, \end{cases} \quad (2)$$

Using the variance of Murthy's estimator [25],

$$V(\bar{y}_I) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \left( 1 - \sum_{s \in \{i,j\}} \frac{\Pr(s|i)\Pr(s|j)}{\Pr(s)} \right) \left( \frac{y_i}{p_i} - \frac{y_j}{p_j} \right)^2 p_i p_j, \text{ where } \Pr(s)$$

is the probability of obtaining sample  $s$ ,  $\Pr(s|i)$  is the conditional probability of obtaining  $s$  given the  $i^{\text{th}}$  unit was selected first,  $p_i$  is the selection probability of unit  $i$  in the first draw and  $S^2 = (N-1)^{-1} \sum_{i=1}^N (y_i - \mu_y)^2$ .

For any sequential sampling design, the sampling procedure may not be able to complete in some situations. This may happen in an inverse sampling design because  $r$  is too large or the population contains very few units satisfying condition  $C$ . General inverse sampling was proposed in [21]. Beginning with a sample size  $n_I$  selected by SRSWOR, we stop further sampling if at least  $r$  units from  $U_M$  are selected. Otherwise, sampling continues until  $r$  units from  $U_M$  are obtained, but a limit is placed on the final sample size, such that  $n_F = N$ . Then an unbiased estimator of  $\mu_y$  is:

$$\bar{y}_{GIS} = \begin{cases} \frac{1}{n_I} \sum_{i=1}^{n_I} y_i & \text{if } [A_1], \\ \frac{1}{N} (M\bar{y}_M + (N-M)\bar{y}_{M'}) & \text{if } [A_2], \\ \frac{1}{n_F} \sum_{i=1}^{n_F} y_i & \text{if } [A_3], \end{cases} \quad (3)$$

where  $[A_1] = \{n_T = n_I\}$ ,

$[A_2] = \{n_I < n_T < n_F \text{ or } n_T = n_F \ \& \ |S_M| = r\}$  and

$[A_3] = \{n_T = n_F \ \& \ |S_M| < r\}$ .

The variance of  $\bar{y}_{GIS}$  is:

$$V(\bar{y}_{GIS}) = \begin{cases} \frac{(N-n_I)S^2}{N n_I} & \text{if } [A_1], \\ V(\bar{y}_I) & \text{if } [A_2], \\ \frac{(N-n_F)S^2}{N n_F} & \text{if } [A_3], \end{cases} \quad (4)$$

The unbiased estimator of the variance of  $\bar{y}_{GIS}$  is:

$$\hat{V}(\bar{y}_{GIS}) = \begin{cases} \frac{(N-n_I)s_I^2}{N n_I} & \text{if } [A_1], \\ \hat{V}(\bar{y}_I) & \text{if } [A_2], \\ \frac{(N-n_F)s_F^2}{N n_F} & \text{if } [A_3], \end{cases} \quad (5)$$

$$\hat{V}(\bar{y}_I) = \frac{1}{N^2} \sum_{i=1}^n \sum_{j<i}^n \left( \frac{\Pr(s|i, j)}{\Pr(s)} - \frac{\Pr(s|i)\Pr(s|j)}{\Pr(s)^2} \right) \left( \frac{y_i}{p_i} - \frac{y_j}{p_j} \right)^2 p_i p_j$$

Here,

where  $\Pr(s|i, j)$  is the probability of obtaining sample  $s$ , given that units  $i$  and  $j$  were selected (in either order) on the first two draws.  $s_I^2 = (n_I - 1)^{-1} \sum_{i=1}^{n_I} (y_i - \bar{y}_I)^2$ ,  $\bar{y}_I = n_I^{-1} \sum_{i=1}^{n_I} y_i$  and  $s_F^2 = (n_F - 1)^{-1} \sum_{i=1}^{n_F} (y_i - \bar{y}_F)^2$ ,  $\bar{y}_F = n_F^{-1} \sum_{i=1}^{n_F} y_i$ .

### 3. General Inverse ACS

The integration of general inverse sampling with ACS is referred to as general inverse ACS. Under this design, if an initially selected unit satisfies the pre-specified condition  $C$ , the entire network to which that unit belongs is included in the sample. Consequently, the final sample comprises the initial units selected through general inverse sampling as well as all units that are members of the networks associated with those initial units. This also includes the edge units of these networks that sequentially satisfy condition  $C$ .

An estimator for the general inverse design with ACS [21]:

$$\bar{y}_{GIS,A} = \begin{cases} \frac{1}{n_I} \sum_{i=1}^{n_I} \bar{y}_i^* & \text{if } [A_1], \\ \frac{1}{N} \left( \hat{M} \bar{y}_M^* + (N - \hat{M}) \bar{y}_{M'}^* \right) & \text{if } [A_2], \\ \frac{1}{n_F} \sum_{i=1}^{n_F} \bar{y}_i^* & \text{if } [A_3], \end{cases} \quad (6)$$

where  $\bar{y}_i^*$  is the average of the variable of interest in the network that includes unit  $i$  of the initial

sample, defined as  $\bar{y}_i^* = \frac{1}{m_i} \sum_{j \in \psi_i} y_j$ , where  $\psi_i$  denotes the network that includes unit  $i$ , and  $m_i$  is the number of units in that network.

The variance of  $\bar{y}_{GIS,A}$  is:

$$V(\bar{y}_{GIS,A}) = \begin{cases} \frac{(N-n_I) S^{*2}}{N n_I} & \text{if } [A_1], \\ V(\bar{y}_I^*) & \text{if } [A_2], \\ \frac{(N-n_F) S^{*2}}{N n_F} & \text{if } [A_3], \end{cases} \quad (7)$$

$$V(\bar{y}_I^*) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j<i}^N \left( 1 - \sum_{s \ni i, j} \frac{\Pr(s|i)\Pr(s|j)}{\Pr(s)} \right) \left( \frac{\bar{y}_i^*}{p_i} - \frac{\bar{y}_j^*}{p_j} \right)^2 p_i p_j \quad \text{and}$$

$$S^{*2} = (N-1)^{-1} \sum_{i=1}^N (\bar{y}_i^* - \mu_y)^2$$

The unbiased estimator of the variance of  $\bar{y}_{GIS,A}$  is:

$$\hat{V}(\bar{y}_{GIS,A}) = \begin{cases} \frac{(N-n_I) s_I^{*2}}{N n_I} & \text{if } [A_1], \\ \hat{V}(\bar{y}_I^*) & \text{if } [A_2], \\ \frac{(N-n_F) s_F^{*2}}{N n_F} & \text{if } [A_3], \end{cases} \quad (8)$$

$$\hat{V}(\bar{y}_I^*) = \frac{1}{N^2} \sum_{i=1}^n \sum_{j<i}^n \left( \frac{\Pr(s|i, j)}{\Pr(s)} - \frac{\Pr(s|i)\Pr(s|j)}{\Pr(s)^2} \right) \left( \frac{\bar{y}_i^*}{p_i} - \frac{\bar{y}_j^*}{p_j} \right)^2 p_i p_j$$

where

$$s_I^{*2} = (n_I - 1)^{-1} \sum_{i=1}^{n_I} (\bar{y}_i^* - \bar{y}_I^*)^2, \quad \bar{y}_I^* = n_I^{-1} \sum_{i=1}^{n_I} \bar{y}_i^* \quad \text{and} \quad s_F^{*2} = (n_F - 1)^{-1} \sum_{i=1}^{n_F} (\bar{y}_i^* - \bar{y}_F^*)^2,$$

$$\bar{y}_F^* = n_F^{-1} \sum_{i=1}^{n_F} \bar{y}_i^*$$

#### 4. Proposed Estimator in General Inverse ACS

Motivated by [24], the estimator  $\bar{x}_{GIS,A}$  can be calculated similar to  $\bar{y}_{GIS,A}$  by using the  $x$ -variable. The modified regression estimator in general inverse ACS is

$$\bar{y}_{RGIS,A} = \bar{y}_{GIS,A} + B^* (\mu_x - \bar{x}_{GIS,A}), \quad \text{where} \quad B^* = \frac{\sum_{i=1}^N \bar{x}_i^* \bar{y}_i^* - N \mu_y \mu_x}{\sum_{i=1}^N \bar{x}_i^{*2} - N \mu_x^2} \quad \text{is estimated by}$$

$$b^* = \frac{\hat{T}_{xy}^* - N \bar{y}_{GIS,A} \bar{x}_{GIS,A}}{\hat{T}_{xx}^* - N \bar{x}_{GIS,A}^2}$$

$$\hat{T}_{xy}^* = \begin{cases} \frac{N}{n_I} \sum_{i=1}^{n_I} \bar{x}_i^* \bar{y}_i^* & \text{if } [A_1], \\ \hat{M}(\bar{xy}^*)_M + (N - \hat{M})(\bar{xy}^*)_{M'} & \text{if } [A_2], \\ \frac{N}{n_F} \sum_{i=1}^{n_F} \bar{x}_i^* \bar{y}_i^* & \text{if } [A_3], \end{cases} \quad (9)$$

where  $(\bar{xy}^*)_M = r^{-1} \sum_{i \in S_M} \bar{x}_i^* \bar{y}_i^*$  and  $(\bar{xy}^*)_{M'} = (n_T - r)^{-1} \sum_{i \in S_{M'}} \bar{x}_i^* \bar{y}_i^*$ .

$$\hat{T}_{xx}^* = \begin{cases} \frac{N}{n_I} \sum_{i=1}^{n_I} \bar{x}_i^{*2} & \text{if } [A_1], \\ \hat{M}(\bar{xx}^*)_M + (N - \hat{M})(\bar{xx}^*)_{M'} & \text{if } [A_2], \\ \frac{N}{n_F} \sum_{i=1}^{n_F} \bar{x}_i^{*2} & \text{if } [A_3], \end{cases}$$

(10)

where  $(\bar{xx}^*)_M = r^{-1} \sum_{i \in S_M} \bar{x}_i^{*2}$  and  $(\bar{xx}^*)_{M'} = (n_T - r)^{-1} \sum_{i \in S_{M'}} \bar{x}_i^{*2}$ .

The bias of  $\bar{y}_{RGIS,A}$  is

$$\begin{aligned} \text{Bias}(\bar{y}_{RGIS,A}) &= E(\bar{y}_{RGIS,A} - \mu_y) \\ &= -E[b^*(\bar{x}_{GIS,A} - \mu_x)] \\ &= -COV(b^*, \bar{x}_{GIS,A}) \end{aligned}$$

From  $\bar{y}_{RGIS,A} = \bar{y}_{GIS,A} + b^*(\mu_x - \bar{x}_{GIS,A}) = (\bar{y}_{GIS,A} - b^*\bar{x}_{GIS,A}) + b\mu_x$ , let  $z_i^* = \bar{y}_i^* - B^*\bar{x}_i^*$ , and let  $B^*$  be estimated by  $b^*$ . Therefore,  $\bar{y}_{RGIS,A} = \bar{z}_{GIS,A} + b^*\mu_x$ .

The variance of  $\bar{y}_{RGIS,A}$  is:

$$V(\bar{y}_{RGIS,A}) = \begin{cases} \frac{(N - n_I) S_z^{*2}}{N n_I} & \text{if } [A_1], \\ V(\bar{y}_{RI}^*) & \text{if } [A_2], \\ \frac{(N - n_F) S_z^{*2}}{N n_F} & \text{if } [A_3], \end{cases}$$

(11)

Using the variance of Murthy's estimator,

$$V(\bar{y}_{RI}^*) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j < i}^N \left( 1 - \sum_{s \neq i, j} \frac{\Pr(s|i)\Pr(s|j)}{\Pr(s)} \right) \left( \frac{z_i^*}{p_i} - \frac{z_j^*}{p_j} \right)^2 p_i p_j \quad \text{and} \quad S_z^{*2} = \frac{1}{N-1} \sum_{i=1}^N (z_i^* - \mu_y)^2$$

The estimator of the variance of  $\bar{y}_{RGIS,A}$  is:

$$\hat{V}(\bar{y}_{GIS,A}) = \begin{cases} \frac{(N - n_I) s_{zI}^{*2}}{N n_I} & \text{if } [A_1], \\ \hat{V}(\bar{y}_{RI}^*) & \text{if } [A_2], \\ \frac{(N - n_F) s_{zF}^{*2}}{N n_F} & \text{if } [A_3], \end{cases}$$

(12)

where  $\hat{V}(\bar{y}_{RI}^*) = \frac{1}{N^2} \sum_{i=1}^n \sum_{j < i}^n \left( \frac{\Pr(s|i, j)}{\Pr(s)} - \frac{\Pr(s|i)\Pr(s|j)}{\Pr(s)^2} \right) \left( \frac{z_i^*}{p_i} - \frac{z_j^*}{p_j} \right)^2 p_i p_j$ , and

$$\begin{aligned} \hat{V}(\bar{y}_{RI}^*) &= \left( \frac{\hat{M}}{N} \right)^2 \left( \frac{(N - n_T + 1)(n_T r - n_T - r) - N(n_T - 2)}{N(n_T - 2)(r - 1)} \right) \frac{s_M^{*2}}{r} \\ &\quad + \frac{1}{N^2} \hat{V}(\hat{M}) (\bar{z}_M^* - \bar{z}_{M'}^*)^2 + \left( \frac{(N - n_T + 1)(n_T - r - 1)}{N(n_T - 1)(n_T - 2)} \right) s_{M'}^{*2} \end{aligned}$$

$$s_M^{*2} = (r-1)^{-1} \sum_{i \in S_M} (z_i^* - \bar{z}_M^*)^2, \quad \bar{z}_M^* = r^{-1} \sum_{i \in M} z_i^*,$$

$$s_{M'}^{*2} = (n_T - r - 1)^{-1} \sum_{i \in S_{M'}} (z_i^* - \bar{z}_{M'}^*)^2, \quad \bar{z}_{M'}^* = (n_T - r)^{-1} \sum_{i \in M'} z_i^*,$$

$$s_{z_I}^{*2} = (n_I - 1)^{-1} \sum_{i=1}^{n_I} (z_i^* - \bar{z}_I^*)^2, \quad \bar{z}_I^* = n_I^{-1} \sum_{i=1}^{n_I} z_i^* \quad \text{and}$$

$$s_{z_F}^{*2} = (n_F - 1)^{-1} \sum_{i=1}^{n_F} (z_i^* - \bar{z}_F^*)^2, \quad \bar{z}_F^* = n_F^{-1} \sum_{i=1}^{n_F} z_i^*.$$

### 5. Simulation Studies and Discussion

The population of the auxiliary variable  $X$  and the variable of interest  $Y$  was generated using a Poisson cluster process [26]. The study area was subdivided into 400 square units (20 rows\*20 columns). The population mean of  $Y$ -values is 59.7375, and the correlation coefficient between the  $Y$ -data and  $X$ -data is 0.9973. The criterion for adding units to the sample were defined as  $C = \{y : y > 0\}$

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	0
0	0	0	0	0	0	0	0	0	0	0	0	34	108	35	8	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	11	293	1,022	375	7	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	2	308	1,182	427	7	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	44	103	40	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	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	16	33	8	0	0	0	0	0	5	0	0	0	0	0
0	0	0	0	0	21	434	645	123	0	0	0	19	67	48	16	17	16	0	0
0	0	0	0	42	746	1,230	208	0	0	4	277	818	306	193	657	337	16	0	0
0	0	0	0	19	177	243	26	0	0	2	362	1,306	599	345	1,371	558	18	0	0
0	0	0	0	0	5	4	0	0	0	5	38	222	84	91	207	133	8	0	0
4	0	0	0	0	0	0	0	3	7	0	0	0	0	0	0	0	25	0	0
0	0	0	0	4	2	0	0	0	0	0	0	0	0	32	355	735	120	3	0
0	0	0	6	179	383	97	0	0	0	0	0	0	0	50	727	1,241	245	13	0
0	0	0	11	576	1,357	287	8	0	0	0	0	0	0	130	270	62	0	0	0
0	0	0	0	181	517	126	3	0	0	0	0	0	0	0	0	0	0	0	0

Figure 1. The population of the variable of interest ( $Y$ ).

The shaded regions in different colors denote distinct networks.

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
0	0	0	0	0	0	0	0	0	0	0	0	8	29	9	2	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	3	74	257	98	3	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	1	73	290	109	2	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	10	26	10	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	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	4	8	2	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	5	108	160	31	0	0	0	4	17	13	4	4	4	0	0
0	0	0	0	11	189	310	52	0	0	1	55	205	76	48	160	84	5	0	0
0	0	0	0	5	45	60	7	0	0	1	87	323	147	87	337	138	5	0	0
0	0	0	0	0	1	1	0	0	0	1	9	57	21	22	55	37	2	0	0
1	0	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0	1	0	0
0	0	0	0	2	1	0	0	0	0	0	0	0	7	91	177	31	1	0	0
0	0	0	2	44	96	25	0	0	0	0	0	0	13	185	311	60	3	0	0
0	0	0	3	148	337	72	2	0	0	0	0	0	0	30	66	14	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

**Figure 2.** The population of the auxiliary variable ( $x$ ).

The position of the network is aligned with the data  $y$ .

A total of 10,000 iterations were performed for each estimator.  $E(n_T)$  is the expected sequential sample size from general inverse sampling and  $E(v)$  is the expected final sample size, respectively.

The estimated variance of the estimators is defined as:

$$\hat{V}(\bar{y}_{GIS,A}) = \frac{1}{10,000-1} \sum_{i=1}^{10,000} \left[ (\bar{y}_{GIS,A})_i - \frac{\sum_{i=1}^{10,000} (\bar{y}_{GIS,A})_i}{10,000} \right]^2$$

$$\hat{V}(\bar{y}_{RGIS,A}) = \frac{1}{10,000-1} \sum_{i=1}^{10,000} \left[ (\bar{y}_{RGIS,A})_i - \frac{\sum_{i=1}^{10,000} (\bar{y}_{RGIS,A})_i}{10,000} \right]^2.$$

The relative efficiency of the proposed estimator, compared with  $\bar{y}_{GIS-A}$ , is defined as

$$RE(\bar{y}_{RGIS,A}) = \frac{\hat{V}(\bar{y}_{GIS,A})}{\hat{V}(\bar{y}_{RGIS,A})}.$$

Since the proposed estimator  $(\bar{y}_{RGIS,A})$  is a biased estimator, the estimated absolute relative bias

$$ARB(\bar{y}_{RGIS,A}) = \frac{1}{10,000} \sum_{i=1}^{10,000} \left| \frac{(\bar{y}_{RGIS,A})_i - \mu_y}{\mu_y} \right|$$

is defined as

### Discussion

Based on the data used in this study, the variable of interest and the auxiliary variable exhibit a linear relationship in the same direction. The results of the study are as follows:

From Table 1, since the proposed estimator  $(\bar{y}_{RGIS,A})$  is a biased estimator, the study found that as  $r$  and  $n_I$  increase, the estimated absolute relative bias of the proposed estimator decreases.

From Table 2, the study of the estimated variance of the estimators showed that as  $r$  and  $n_I$  increase, the estimated variance of the estimators decreases. When comparing the estimated variance between the proposed estimator that uses auxiliary variable  $(\bar{y}_{RGIS,A})$  and the estimator of the mean of the variable of interest without using auxiliary variable  $(\bar{y}_{GIS,A})$ , it was found that  $(\bar{y}_{RGIS,A})$  has a lower estimated variance than  $(\bar{y}_{GIS,A})$  in all cases. Moreover, the relative efficiency of the proposed estimator is greater than one in all cases, indicating that  $(\bar{y}_{RGIS,A})$  is more efficient than  $(\bar{y}_{GIS,A})$ .

**Table 1.** The estimated absolute relative bias of the proposed estimator for the population mean of the variable of interest.

$r$	$n_I$	$E(n_T)$	$E(v)$	$ARB(\bar{y}_{RGIS,A})$
2	5	8.7335	57.0359	0.37475
2	10	11.2945	65.1829	0.19846
2	15	15.4288	75.4269	0.06664
3	5	12.2907	69.8711	0.21408
3	10	13.4347	72.8004	0.20585
3	15	16.5289	79.6913	0.16196

3	20	20.4772	87.5785	0.04360
4	10	16.6089	82.3383	0.14986
4	15	18.0372	84.5381	0.15376
4	20	21.3896	91.4283	0.12636
5	15	21.3911	91.4934	0.13535
5	20	23.6009	95.6987	0.13861
5	25	26.2153	100.5247	0.09379
5	30	30.5074	106.9697	0.04137
10	30	41.5436	122.0440	0.09839
10	40	44.2324	124.5140	0.08867
10	50	44.2324	130.6097	0.05299

**Table 2.** The estimated variance of the estimators for the population mean of the variable of interest and the relative efficiency of the proposed estimator compared with  $\bar{y}_{GIS,A}$ .

$r$	$n_I$	$E(n_T)$	$E(v)$	$\hat{V}(\bar{y}_{GIS,A})$	$\hat{V}(\bar{y}_{RGIS,A})$	$RE(\bar{y}_{RGIS,A})$
2	5	8.7335	57.0359	2,242.0274	1,359.3310	1.6494
2	10	11.2945	65.1829	1,174.3768	663.9508	1.7688
2	15	15.4288	75.4269	730.6367	579.0399	1.2618
3	5	12.2907	69.8711	1,772.1519	1,287.5967	1.3763
3	10	13.4347	72.8004	863.4391	389.3139	2.2178
3	15	16.5289	79.6913	795.5099	426.5605	1.8649
3	20	20.4772	87.5785	526.7130	406.0654	1.2971
4	10	16.6089	82.3383	742.8746	422.3135	1.7591
4	15	18.0372	84.5381	544.0685	232.7903	2.3372
4	20	21.3896	91.4283	536.4767	285.0840	1.8818
5	15	21.3911	91.4934	516.0823	241.1235	2.1403
5	20	23.6009	95.6987	524.0002	228.0293	2.2980
5	25	26.2153	100.5247	398.3312	219.3141	1.8163
5	30	30.5074	106.9697	344.7272	228.7729	1.5069
10	30	41.5436	122.0440	339.7744	167.0369	2.0341
10	40	44.2324	124.5140	194.7799	72.3611	2.6918
10	50	44.2324	130.6097	152.7576	93.7495	1.6294

## 6. Conclusions

General inverse sampling was incorporated into ACS in [21], and this design is referred to as general inverse ACS. In this paper, an estimator of the mean of the variable of interest in general inverse ACS is developed using auxiliary variable, where the auxiliary variable has a linear relationship with the variable of interest. The study found that the proposed estimator is more efficient than the estimator of the mean of the variable of interest that does not use auxiliary variable, across all investigated scenarios.

In this study, the proposed estimator was motivated by the work in [24], which introduced a regression-type estimator utilizing information from a single auxiliary variable. Therefore, future research could focus on developing estimators that use information from other auxiliary variables, such as the coefficient of variation of the auxiliary variable or the correlation coefficient between the auxiliary variable and the variable of interest as well as the use of more than one auxiliary variable.

**Author Contributions:** Conceptualization, N.C. and P.G.; methodology, S.W. and N.C.; software, N.C.; validation, M.C. and C.B.; formal analysis, N.C. and M.C.; investigation, S.W. and P.G.; resources, N.C. and C.B.; data curation, N.C. and C.B.; writing—original draft preparation, N.C. and M.C.; writing—review and editing, P.G. and S.W.; visualization, P.G. and C.B.; supervision, N.C. and P.G.; project administration, N.C.; funding acquisition, N.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research project was financially supported by Mahasarakham University.

**Data Availability Statement:** The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

**Conflicts of Interest:** The authors have no conflicts of interest to declare that are relevant to the content of this article.

## References

1. Thompson, S.K. Adaptive cluster sampling. *J. Am. Statist. Assoc.* **1990**, *85*, 1050–1059. <https://doi.org/10.2307/2289601>.
2. Magnussen, S.; Kurz, W.; Leckie, D.G.; Paradine, D. Adaptive cluster sampling for estimation of deforestation rates. *Eur. J. For. Res.* **2005**, *124*, 207–220. <https://doi.org/10.1007/s10342-005-0074-6>.
3. Noon, B.R.; Ishwar, N.M.; Vasudevan, K. Efficiency of adaptive cluster and random sampling in detecting terrestrial herpetofauna in a tropical rainforest. *Wildl. Soc. Bull.* **2006**, *34*, 59–68. [https://doi.org/10.2193/0091-7648\(2006\)34\[59:EOACAR\]2.0.CO;2](https://doi.org/10.2193/0091-7648(2006)34[59:EOACAR]2.0.CO;2).
4. Sullivan, W.P.; Morrison, B.J.; Beamish, F.W.H. Adaptive cluster sampling: Estimating density of spatially autocorrelated larvae of the sea lamprey with improved precision. *J. Great Lakes Res.* **2008**, *34*, 86–97. [https://doi.org/10.3394/0380-1330\(2008\)34\[86:ACSEDO\]2.0.CO;2](https://doi.org/10.3394/0380-1330(2008)34[86:ACSEDO]2.0.CO;2).
5. Smith, D.R.; Villella, R.F.; Lemarié, D.P. Application of adaptive cluster sampling to low-density populations of freshwater mussels. *Environ. Ecol. Stat.* **2003**, *10*, 7–15. <https://doi.org/10.1023/A:1021956617984>.
6. Conners, M.E.; Schwager, S.J. The use of adaptive cluster sampling for hydroacoustic surveys. *ICES J. Mar. Sci.* **2002**, *59*, 1314–1325. <https://doi.org/10.1006/jmsc.2002.1306>.
7. Olayiwola, O.M.; Ajayi, A.O.; Onifade, O.C.; Wale-Orojo, O.; Ajibade, B. Adaptive cluster sampling with model based approach for estimating total number of Hidden COVID-19 carriers in Nigeria. *Stat. J. IAOS* **2020**, *36*, 103–109. <https://doi.org/10.3233/SJI-200718>.
8. Chandra, G.; Tiwari, N.; Nautiyal, R. Adaptive cluster sampling-based design for estimating COVID-19 cases with random samples. *Curr. Sci.* **2021**, *120*, 1204–1210. <https://doi.org/10.18520/cs/v120/i7/1202-1210>.
9. Stehlík, M.; Kiseřák, J.; Dinamarca, A.; Alvarado, E.; Plaza, F.; Medina, F.A.; Stehlíková, S.; Marek, J.; Venegas, B.; Gajdoš, A.; et al. REDACS: Regional emergency-driven adaptive cluster sampling for effective COVID-19 management. *Stoch. Anal. Appl.* **2022**, *41*, 474–508. <https://doi.org/10.1080/07362994.2022.2033126>.
10. Hwang, J.; Bose, N.; Fan, S. AUV adaptive sampling methods: A Review. *Appl. Sci.* **2019**, *9*, 3145. <https://doi.org/10.3390/app9153145>.
11. Giouroukis, D.; Dadiani, A.; Traub, J.; Zeuch, S.; Markl, V. A survey of adaptive sampling and filtering algorithms for the internet of things. In Proceedings of the 14th ACM International Conference on Distributed and Event- Based Systems, Montreal, QC, Canada, 13–17 July 2020; pp. 27–38. <https://doi.org/10.1145/3401025.3403777>.
12. Chao, C.T. Ratio estimation on adaptive cluster sampling. *J. Chin. Stat. Assoc.* **2004**, *42*, 307–327. <https://doi.org/10.29973/JCSA.200409.0006>.
13. Dryver, A.L.; Chao, C.T. Ratio estimators in adaptive cluster sampling. *Environmetric* **2007**, *18*, 607–620. <https://doi.org/10.1002/env.838>.
14. Chutiman, N.; Kumphon, B. Ratio estimator using two auxiliary variables for adaptive cluster sampling. *Thail. Stat.* **2008**, *6*, 241–256.
15. Chutiman, N. Adaptive cluster sampling using auxiliary variable. *J. Math. Stat.* **2013**, *9*, 249–255. <https://doi.org/10.3844/jmssp.2013.249.255>.
16. Yadav, S.K.; Misra, S.; Mishra, S. Efficient estimator for population variance using auxiliary variable. *Am. J. Oper. Res.* **2016**, *6*, 9–15. <https://doi.org/10.5923/j.ajor.20160601.02>.
17. Chaudhry, M.S.; Hanif, M. Generalized exponential-cum-exponential estimator in adaptive cluster sampling. *Pak. J. Stat. Oper. Res.* **2015**, *11*, 553–574. <https://doi.org/10.18187/pjsor.v11i4.1009>.
18. Chaudhry, M.S.; Hanif, M. Generalized difference-cum-exponential estimator in adaptive cluster sampling. *Pak. J. Stat.* **2017**, *33*, 335–367.

19. Bhat, A.A.; Sharma, M.; Shah, M.; Bhat, M. Generalized ratio type estimator under adaptive cluster sampling. *J. Sci. Res.* **2023**, *67*, 46–51. <https://doi.org/10.37398/JSR.2023.670307>.
20. Christman, M.C.; Lan, F. Inverse Adaptive Cluster Sampling. *Biometrics*, **2001**, *57*, 1096–1105. <https://doi.org/10.1111/j.0006-341X.2001.01096.x>
21. Salehi, M.; Seber, G.A.F. A general inverse sampling scheme and its application to adaptive cluster sampling. *Aust. N. Z. J. Stat.* **2004**, *46*, 483 – 494. <https://doi.org/10.1111/j.1467-842X.2004.00345.x>
22. Pochai, N. An Improved the Estimator in Inverse Adaptive Cluster Sampling. *Thail. Stat.* **2008**, *6*, 15–26. <https://ph02.tci-thaijo.org/index.php/thaistat/article/view/34329>
23. Sangngam, P. Unequal Probability Inverse Adaptive Cluster Sampling. *Chiang Mai J. Sci.* **2013**, *40*, 736-742.
24. Moradi, M.; Salehi, M.; Brown, J.A.; Karimi, N. (2011). Regression estimator under inverse sampling to estimate arsenic contamination. *Environmetrics.* **2011**, *22*, 894-900. <https://doi.org/10.1002/env.1116>.
25. Salehi, M.; Seber, G.A.F. (2001). Theory & Methods: A New Proof of Murthy's Estimator which Applies to Sequential Sampling. *Aust. N. Z. J. Stat.* **2001**, *43*, 281–286.
26. Subzar, M.; Alqurashi, T.; Chandawat, D.; Tamboli, S.; Raja, T. A.; Attri, A.K.; Wani, S.A. Generalized robust regression techniques and adaptive cluster sampling for efficient estimation of population mean in case of rare and clustered populations. *Sci Rep.* **2025**, *15*, 2069. <https://doi.org/10.1038/s41598-025-85328-0>.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.