Estimation of Bounded Populations and Carrying Capacity with the Logistic Model

Lamin Kabareh[1*], Thomas Mageto[2]

[1]Pan African University Institute for Basic Sciences, Technology and Innovation (PAUSTI), Nairobi, Kenya
[2]Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya
Email: *lkabareh@yahoo.com

Abstract

Estimation of bounded populations and carrying capacity in the presence of a sample frame is considered. Models based on Logistic model are proposed. Like the existing estimators, this estimation technique deals with initial condition and is based on yearly population totals in order to fit in a model within a given period of time in this study. The proposed Logistic model technique has shown to be efficient especially with large data. The empirical study indicated that the Logistic model is efficient and can estimate properly even in the presence of outliers.

Keywords

Logistic Model, Estimation, Bounded Population, Carrying Capacity, Sample Frame, Empirical Study and Outliers

1. Introduction

Sample surveys are widely used as a cost effective apparatus of data collection and for making valid inference about population parameters. Government bureaus and organizations use such methods to obtain the current information. The foremost aim of a statistician in a sample survey is to obtain information about the population by deriving reliable estimates of unknown population parameters.

This study is using estimation techniques to estimate the bounded population and carrying capacity called the Logistic model that do not require any choice of step size as in the case of local polynomial regression estimator or have to be restricted a fix behavior, instead we allow the data to reveal its nature. The logistic model is use for data fitting. The logistic equation was introduced (around 1840) by the Belgian mathematician and demographer P.F. Verhulst as a possible

model for human population growth [1].

Under simple random sampling (SRS) without replacement design, [2] proposed an exactly unbiased estimator for θ_{yx}. The proposed estimator is given by

$$\hat{\theta}_{HR} = \bar{r}_s + \frac{n(N-1)}{N(n-1)\bar{x}_u}\left(\bar{y}_s - \bar{r}_s\bar{x}_s\right) \tag{1.1}$$

where, $\bar{y}_s = \sum_{i \in s}\frac{y_i}{n}$, $\bar{r}_s = \sum_{i \in s}\frac{r_i}{n}$, $r_i = \frac{y_i}{x_i}$, $\bar{x}_s = \sum_{i \in s}\frac{x_i}{n}$, $\bar{x}_u = \frac{t_x}{N}$, the population ratio $\theta_{yx} = \frac{t_y}{t_x}$, where $t_y = \sum_{i \in U} y_i$ *be the population total for the variable Y,* $t_x = \sum_{i \in U} x_i$ *be the population total for the variable X and U of N units indexed by the set* $\{1, 2, \cdots, N\}$ *a finite population.* This estimator can be rewritten under general sampling design $p(\cdot)$. In this case, this estimator is no longer unbiased but still with negligible bias [3].

Under general sampling design, [4] proposed an estimator for estimating the population ratio θ_{yx}. This estimator, has negligible relative bias especially for small sample sizes n and approaches zero with increasing n. Under SRS, and based on simulation results, the performance of this estimator is better than that of (1.1). Their estimator is defined by

$$\hat{\theta}_{JM} = \bar{r}_s + \frac{1}{\bar{x}_s}\left(\bar{y}_s - \bar{r}_s\bar{x}_s\right) \tag{1.2}$$

Define π_i, the first order inclusion probability, by

$$\pi_i = P_r\left(i^{\text{th}} \text{ element} \in s\right) = \sum_{i,j \in s} P(s) \tag{1.3}$$

For $i \neq j$, the second order inclusion probability is defined by

$$\pi_{ij} = P_r\left(i^{\text{th}} \text{ and } j^{\text{th}} \text{ elements} \in s\right) = \sum_{i,j \in s} P(s) \tag{1.4}$$

The [5] estimator of the population total $t_y = \sum_{i \in U} y_i$ is defined by

$$\hat{t}_{y\pi} = \sum_{i \in U} y_i \frac{I_{\{i \in s\}}}{\pi_i} \tag{1.5}$$

where $I_{\{i \in s\}}$ is one if $i \in s$ and zero otherwise. Further,

$$\bar{y}_s = \frac{1}{N}\hat{t}_{y\pi} \tag{1.6}$$

can be used to estimate the population mean $\bar{y}_u = \frac{1}{N}t_y$. It can be noted that $\hat{t}_{y\pi}$ and $\bar{y}_s$ are unbiased estimators for t_y, and $\bar{y}_u$ respectively. However, $\hat{t}_{y\pi}$ and $\bar{y}_s$ do not use the availability of auxiliary variables in the study. In similar way,

$$\bar{x}_s = \frac{1}{N}\hat{t}_{x\pi} \quad \text{and} \quad \bar{r}_s = \frac{1}{N}\hat{t}_{r\pi} \tag{1.7}$$

are unbiased estimators for $\bar{x}_u$ and $\bar{r}_u$ respectively. Where $\bar{x}_s$ is the sample mean of the inclusion probability of the auxiliary variable.

The availability of more than one auxiliary variable issued in literature for estimating the finite population total t_y, or finite population mean $\overline{y}_u$.

Under SRS, [6] was the first one who deals with the problem of estimating the population mean using more than one auxiliary variables. His estimator is given by

$$\hat{\overline{y}}_u = \sum_{i=1}^{P} w_i \overline{x}_{iu} \hat{\theta}_{yx}$$ (1.8)

where p is the number of the auxiliary variables, $\hat{\theta}_{yx_i} = \dfrac{\overline{y}_s}{\overline{x}_{is}}$ w_i is the weight of

the ith auxiliary variable such that $\sum_{i=1}^{P} w_i = 1$ $\overline{y}_s$ is the sample mean of Y and $\overline{x}_{iu}, \overline{x}_{is}$ are the population mean and the sample mean of X, respectively, for $i = 1, \cdots, p$. [7] proposed the following estimator

$$\hat{\overline{y}}_u = \overline{y}_s \left(w_1 \frac{\overline{x}_{1u}}{\overline{x}_{1s}} + w_2 \frac{\overline{x}_{2u}}{\overline{x}_{2s}} \right)$$ (1.9)

for estimating the population mean $\overline{y}_u$, $w_1 + w_2 = 1$.

[8] studied the general form of (1.9). They proposed two classes of estimators using two auxiliary variables to estimate the population mean for the variable of interest Y.

[9] suggested a new multivariate ratio estimator using the regression estimator instead of $\overline{y}_s$ which used in (1.9). Their estimator is given by

$$\overline{y}_{pr} = \sum_{i=1}^{2} w_i \frac{\overline{y}_s + b_i (\overline{x}_{iu} - \overline{x}_{is})}{\overline{x}_{is}} \overline{x}_{iu}$$ (1.20)

where b_i, $i = 1,2$ are the regression coefficients. Based on the mean squares error (*MSE*), they found that their estimator is more efficient than (1.9) when

$$MSE\left(\overline{y}_{pr}\right) < MSE\left(\overline{y}_u\right),$$

where $MSE\left(\overline{y}_{pr}\right)$, and $MSE\left(\overline{y}_u\right)$ are defined by Equations (2.4), and (1.2) of Kadilar and Cingi (2004), respectively.

In subsection 2.1 we introduced a general population model that accommodates birth and death rates that are necessarily constant, while subsection 2.2 talked about the asymptotic properties and Section 3.1 talked about the empirical studies. Finally, Section 4.0 drew a conclusion on the study. However, our population $P(t)$ will be a continuous approximation to the actual population, which of course changes only by integral increments—that is, by one birth or death at a time.

Suppose that the population changes only by the occurrence of births and deaths—there is no immigration or emigration from outside the country or environment under consideration. It is customary to track the growth or decline of a population in terms of its birth rate and death rate functions defined as follows:

$B(t)$ is the number of births per unit of population per unit of time at time t;

$D(t)$ is the number of deaths that occur during the time at time t.

Then the numbers of births and deaths that occur during the time interval $[t, t + \Delta t]$ is given (approximately) by:

Births: $B(t) \cdot P(t) \cdot \Delta t$,   Deaths: $D(t) \cdot P(t) \cdot \Delta t$

Hence the change ΔP in the population during the time interval $[t, t + \Delta t]$ of length Δt is

$$\Delta P = \{\text{births}\} - \{\text{deaths}\} \approx B(t) \cdot P(t) \cdot \Delta t - D(t) \cdot P(t) \cdot \Delta t \qquad (1.21)$$

So

$$\frac{\Delta P}{\Delta t} \approx \left[B(t) - D(t) \right] P(t) \qquad (1.22)$$

The error in this approximation should approach zero as $\Delta t \to 0$, so—taking the limit—we get the differential equation

$$\frac{\mathrm{d}P}{\mathrm{d}t} = (B - D) P \qquad (1.23)$$

in which we write $B = B(t)$, $D = D(t)$, and $P = P(t)$ for brevity. Equation (1.22) is the general population equation. If B and D are constants, Equation (1.22) reduces to the natural growth equation with $K = B - D$. But it also includes the possibility that B and D are variable functions of t. The birth and death rates need not be known in advance; they may well depend on the unknown function $P(t)$.

2. Estimation of Bounded Population and Carrying Capacity

This section is purposely considering an estimator that is the logistic model estimate of the bounded population and carrying capacity.

2.1. Proposed Logistic Model

Suppose the birth rate B is a linear decreasing function of the population size P, so that $B = B_0 - B_1 P$, where B_0 and B_1 are positive constants. If the death rate $D = D_0$ remains constant, then Equation (1.22) takes the form

$$\frac{\mathrm{d}P}{\mathrm{d}t} = (B_0 - B_1 P - D_0) P \qquad (1.24)$$

That is,

$$\frac{\mathrm{d}P}{\mathrm{d}t} = aP - bP^2 \qquad (1.25)$$

where $a = B_0 - D_0$ and $b = B_1$

If the coefficients a and b are both positive, then Equation (1.25) is called the logistic equation. For the purpose of relating the behavior of the population $P(t)$ to the values of the parameters in the equation, it is useful to rewrite the logistic equation in the form

$$\frac{\mathrm{d}P}{\mathrm{d}t} = KP(M - P), \ P(0) = P_0 \qquad (1.26)$$

where $K = b$ and $M = \dfrac{a}{b}$ are constants. Solving Equation (1.25) gives,

$$P(t) = \frac{MP_0}{P_0 + (M - P_0)e^{-KMt}} \qquad (1.27)$$

Actual human populations are positive valued. If $P_0 = M$, then (1.27) reduces to the unchanging (constant- valued) "equilibrium population" $P(t) \equiv M$. Otherwise, the behavior of a logistic population depends on whether $0 < P_0 < M$ or $P_0 > M$. If $0 < P_0 < M$, then we see from (1.26) and (1.27) that $P' > 0$ and

$$P(t) = \frac{MP_0}{P_0 + (M - P_0)e^{-KMt}} = \frac{MP_0}{P_0 + \{\text{postive number}\}} < \frac{MP_0}{P_0} = M.$$

However, if $P_0 > M$, then we see from (1.26) and (1.27) that $P' < 0$ and

$$P(t) = \frac{MP_0}{P_0 + (M - P_0)e^{-KMt}} = \frac{MP_0}{P_0 + \{\text{negative number}\}} > \frac{MP_0}{P_0} = M.$$

In either case, the "positive number" or "negative number" in the denominator has absolute value less than P_0 an—because of the exponential factor- approaches 0 as $t \to +\infty$. It follows that

$$\lim_{t \to \infty} P(t) = \frac{MP_0}{P_0 + 0} = M \qquad (1.28)$$

Thus a population that satisfies the logistic equation does not grow without bound. Instead, it approaches the finite limiting population M as $t \to \infty$. The population $P(t)$ steadily increases and approaches M from below if $0 < P_0 < M$, but steadily decreases and approaches M from above if $P_0 > M$. Sometimes M is called the carrying capacity of the environment, considering it to be the maximum population that the environment can support on a long-term basis.

The five census years obtained from a sample frame is shown in Table 1 above. However, we aimed at selecting 1969 population as initial population and fit a model through 1989 and 2009 populations from the table. These sample sizes will be used to estimate the population total in 2019 census using the proposed techniques.

Here, $P_0 = 10942705$ (Initial population)

At $t_2 = 1989$ and $P_2 = 21443636$ we have;

$$\frac{10942705M}{10942705 + (M - 10942705)e^{-20KM}} = 21448774 \qquad (1.29)$$

Table 1. Census Results.

i	YEAR ($X_i = t_i$)	POPULATION TOTAL ($Y_i = P_i$)
0	1969	10,942,705
1	1979	15,327,061
2	1989	21,448,774
3	1999	28,686,607
4	2009	38,610,097

Similarly,

At $t_4 = 2009$ and $P_4 = 38610097$ we have;

$$\frac{10942705M}{10942705 + (M - 10942705)\mathrm{e}^{-40KM}} = 38610097 \qquad (1.30)$$

Solving Equations ((1.29) and (1.30)) simultaneously we have;

$$\{KM = 0.038506784, M = 124433288.8\}$$

$$P(t) = \frac{1.361636772 \times 10^{15}}{10942705 + 113490583.8\mathrm{e}^{-0.038506784t}} \qquad (1.31)$$

2.2. Asymptotic Properties

Theorem: Law of large numbers:

Let $X_1, X_2, \cdots, X_n$ be iid random variables with common expectation $\mu = E(X_i)$. Define $A_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then for any $\propto > 0$, we have $P_r\left[|A_n - \mu| \geq \propto\right] \to 0$ as $n \to \infty$

Proof of Theorem:

Let $Var(X_i) = \sigma^2$ be the common variance of the random variables; we assume that σ^2 is finite. With this (relatively mild) assumption, the Law of Large Numbers (LLN) is an immediate consequence of Chebyshev's inequality. For as we have seen above, $E(A_n) = \mu$ and $Var(A_n) = \frac{\sigma^2}{n}$, so by Chebyshev we have

$$P_r\left[|A_n - \mu| \geq \propto\right] \leq \frac{Var(A_n)}{\propto^2} = \frac{\sigma^2}{n\alpha^2} \to 0 \text{ as } n \to \infty$$

Table 2 above represents the census population from 1969 to 2009 in the eight provinces in Kenya. Successive sample sizes are selected below to show the law of large numbers.

Here, $N = 40$ and $\mu = 2875251$

Table 2. Provinces.

Provinces	1969	1979	1989	1999	2009
Nairobi	509,286	827,775	1,324,570	2,143,254	3,138,369
C. Province	1,675,647	2,345,833	3,116,703	3,724,159	4,383,743
Coast Province	944,082	1,342,794	1,829,191	2,487,264	3,325,307
E. Province	1,907,301	2,719,851	3,768,677	4,631,779	5,668,123
N.E. Province	245,757	373,787	371,391	962,143	2,310,757
Nyanza	2,122,045	2,643,956	3,507,162	4,392,196	5,442,711
R. Valley	2,210,289	3,240,402	4,981,613	6,987,036	10,006,805
W. Province	1,328,298	1,832,663	2,544,329	3,358,776	4,334,202
Total Population	10,942,705	15,327,061	21,443,636	28,686,607	38,610,097

Sample 1: Nairobi (1969 to 2009)

$n = 5$ and $\overline{x}_1 = 1588651$

Sample 2: Nairobi and Central

$n = 10$ and $\overline{x}_2 = 2318934$

Sample 3: Nairobi, Central and Coast

$n = 15$ and $\overline{x}_3 = 1915957$

Sample 4: Nairobi, Central, Coast & Eastern

$n = 20$ and $\overline{x}_4 = 2371755$

Sample 5: Nairobi, Central, Coast, Eastern and N/Eastern

$n = 25$ and $\overline{x}_5 = 2067957$

Sample 6: Nairobi, Central, Coast, Eastern, N/Eastern and Nyanza

$n = 30$ and $\overline{x}_6 = 2326900$

Sample 7: sample 6 and R. Valley

$n = 35$ and $\overline{x}_7 = 2778090$

Remark: We can clearly see the sample mean tending to the population mean as we approach the population total N which is in line with the Law of Large Numbers (LLN)

Therefore, $\lim_{n \to N} \overline{x}_n = \mu$

Comment: This technique can track reasonably well throughout up to a sufficiently large number after which, there is a need to shift the initial condition to where the error margin starts increasing in order to maintain precision.

3. Main Results

Empirical Analysis

Table 3 represents the actual population totals, estimated population totals and their corresponding errors from 1969 to 2009.

4. Conclusion

In this work, the logistic model is very effective especially with the presence of outliers in trying to maintain precision. It can perform well with a sufficiently large sample size. The logistic model can be more efficient in prediction especially where a regression model is ill conditioned.

Table 3. Estimated population and error calculations.

YEAR(t_i)	ACTUAL POPULATION(P_i)	ESTIMATED POPULATION(P_i)	ERROR(t_i)
1969 ($t = 0$)	10,942,705	10,942,705	0
1979 ($t = 10$)	15,327,061	15,444,677	-117,616
1989 ($t = 20$)	21,443,636	21448,774	-5138
1999 ($t = 30$)	28,686,607	29,162,417	-475810
2009 ($t = 40$)	38,610,097	38,610,096	1

$P_{2019}(50) = 49527365$

Acknowledgements

Disclosure of Potential Conflicts of Interest

Authors strongly disclose no conflict of interest with regard to the publication of the paper.

References

[1] Edwards, C.H. and Penney, D.E. (2008) Differential Equations: Computing and Modeling. 4th Edition, 79-92.

[2] Hartley, H. and Ross, A. (1954) Unbiased Ratio Estimates. *Nature*, **174**, 270-271. https://doi.org/10.1038/174270a0

[3] Al-Jararha, J. (2012) Unbiased Ratio Estimation for Finite Populations. LAMBERT Academic Publishing, Germany.

[4] Al-Jararha, J. and Al-Haj, E.M. (2012) A Ratio Estimator Under General Sampling Design. *Austrian Journal of Statistics*, **41**, 105-115. https://doi.org/10.17713/ajs.v41i2.178

[5] Horvitz, D. and Thompson, D. (1952) A Generalization of Sampling without Replacement from a Finite Universe. *Journal of the American Statistical Association*, **47**, 663-685. https://doi.org/10.1080/01621459.1952.10483446

[6] Olkin, I. (1958) Multivariate Ratio Estimation for the Finite Populations. *Biometrika*, **45**, 154-165. https://doi.org/10.1093/biomet/45.1-2.154

[7] Singh, D. and Chaudhary, F. (1986) Theory and Analysis of Sample Survey Design. New Age Publication, New Delhi.

[8] Abu-Dayyeh, W., Ahmad, M., Ahmad, R. and Hassen, A. (2003) Some Estimators of a Finite Population Mean Using Auxiliary Information. *Applied Mathematics and Computations*, **139**, 287-298. https://doi.org/10.1016/S0096-3003(02)00180-7

[9] Kadilar, C. and Cingi, H. (2004) Estimator of a Population Mean Using Two Auxiliary Variables in Simple Random Sampling. *International Mathematical Journal*, **5**, 357-367.