BIBLIOGRAPHY

ALOS, EMILY JUNE C., BAT-OY, ELEONOR B. April 2008. Efficiency of Ratio and Regression Estimators in Estimating Strawberry Production. Benguet, State University, La Trinidad, Benguet.

Adviser: Marycel H. Toyhacao

ABSTRACT

The study was conducted to determine the efficiency of the estimates on strawberry production using ratio and regression estimators.

A total of 195 plots were selected as samples from the total population of 899 plots from the Sariling Sikap area at strawberry farm. The data were summarized, tabulated and analyzed. The total production and total number of plants per area were used as the variables in the estimation.

Being an unbiased estimator, the Simple Random Sampling has the least mean and total production of strawberry. Among the bias estimator, classical ratio estimator is the most precise among the three estimators used in estimating strawberry production with its least coefficient of variability.

Regression estimator is the most consistent estimator and it is considered the most efficient estimator among the estimator used in estimating the strawberry production.

TABLE OF CONTENTS

INTRODUCTION

Background of the Study

A sample survey may be considered as an absolute (enumeration) experiment whose main objective are to obtain estimators of parameters and to derive measures of precision of these estimators. In analytical surveys one of the objectives is to test hypothesis with the use of appropriate estimation procedures. Thus, it is apparent that simple and sound estimation techniques must be developed to provide precise and accurate statistics whether such endeavors are of the absolute or analytical type or both. (Burton T. Onate & Julia Mercedes O. Bader: 1990)

Estimation of the population mean and total sometimes based on a sample of response measurements $y_1, y_2, \ldots. y_n$ are obtained by simple random sampling and stratified random sampling. William Mendenhall (1990) stated that measuring Y and one or more subsidiary variables are used to estimate the mean of the response Y. It is basic to the correlation and provides means for development of a prediction equation relating Y and X by the method of least square.

The basic concept in estimation procedure is to determine the unknown values of the parameters using the sample values. By doing so, many estimators can be utilized. Among of those are using ratio estimators and regression estimators. However, the correct approach can lead to better estimates, that is, less bias is obtained. Ratio estimators and regression estimators, as pointed by Onate and Bader (1990), increased its precision or usually attained by observing an additional variable Z_i, in addition to the original characteristics.

It is most effective when the relationship between the response y and the subsidiary variable x is linear through the origin and the variance of y is proportional to x.

In Regression estimation the technique is for incorporating information on a subsidiary variable, this is better if the relationship between the y's and x's is a straight line not through the origin. It works well when the plot of y and x reveals points by lying uniformly close to a straight line with unit slope.

Ratio estimation usually leads to biased estimators. Thus, a consideration for the magnitude of the bias should be observed. However, for a large sample size (n>30) and he $(\sigma_x/\mu_x) \geq 10$, the bias is negligible. Ratio estimators are unbiased when the relationship between y and x is linear through the origin. As mentioned, we can assume to draw sample from the population, and possibly estimate its average (strawberry production) from y, sample mean. Since there are underlying associations between the mentioned variables, the number of plants per plot can be multiplied from the production of strawberries per plot.

Objective of the Study

Specifically, the study was conducted to determine the efficiency of the estimates on the production of strawberries in Strawberry Farm, La Trinidad, Benguet using the ratio estimates and using regression estimates.

Importance of the Study

The researchers find the study important; it gives information to the concern individuals specially in estimating the production of strawberry in La Trinidad, Benguet using the ratio and regression estimators. Through this study, students would be aware that strawberry production can be estimated or it can be measured using statistical tools by studies and researchers.

The useful and relevant information acquired from the study will encourage future researchers to estimate strawberry production applying the choice of their analysis. This can help them to boost their ability in analyzing given data's from the specific tools they will conduct.

Result of the study will provide the strawberry management knowledge if the ratio and regression estimators can be appropriately applied to the production of strawberries. This will make them aware on the specific materials needed for strawberry production.

This study will serve as a valuable reference to farmers that production of strawberry can be estimated through different estimation procedures. This will also be a basis on how many strawberries will be produced in one plot.

Finally, this study will be an instrument or a benchmark for other researchers who are planning to study on different estimation procedures.

Scope and Delimitation

The main focus of this study was to determine the efficiency of ratio and regression estimators in the production of strawberries in Strawberry Farm at La Trinidad, Benguet.

The data was gathered from the Strawberry Farm of La Trinidad, Benguet. Specifically at the Sariling Sikap Area.

REVIEW OF LITERATURE

Ratio Estimation

The ratio estimation is a different way of estimating population total or mean that is useful in many problems. It involves the use of known population totals for auxiliary variables to improve the weighting from sample values to population estimates. It operates by comparing the survey sample estimate for an auxiliary variable with the known population total for the same variable on the frame. The ratio of the sample estimate of the auxiliary variable to its population total on the frame is used to adjust the sample estimate for the variable of interest.

Ratio method of estimation is frequently used in sample surveys to estimate the population mean of the variable under investigation. Several ratio-type estimates can be formed. All these estimates are unsatisfactory in the sense that they are biased. Hartley and Ross, who proposed an unbiased ratio-type estimator for uni-stage sampling designs, overcame this difficulty. In practice, however, we are generally faced with multi-stage sampling designs. This communication gives a generalized form of Hartley-Ross unbiased ratio-type estimator for multi-stage designs.

This method aims to obtain an increased precision by taking advantage of the correlation between Y_i and X_i .It applies population knowledge of adjustment variable X_i to improve estimation of the population total should be known.

The ratio estimates can be used when X_i is done some other kind of supplementary variables. However, for successful application, the ratio Y_i / X_i should be relatively constant over the population and the population total should be known.

Lohr (1999) enumerated the uses of ratio estimation which include: ratio estimators is simply used to estimate a ratio; to estimate a population total but the population size N is unknown and to adjust estimates from the sample so as to reflect demographic totals.Often,it is used to adjust for no response.

Hsu and Kuo (2000) used the ratio estimation to estimate the recycled and the production values and employment opportunities induced by household waste recycling. Ratio estimation was used in their study because the true population size of recycling plant is unknown; and the audited amount of household waste recycling is known; and the audited amount and recycling amount are highly correlated.

Angel et.Al (2004) evaluated the impact of using ratio estimation in their study, as the estimate census of night dwelling was much closer to the actual value.

M.Gossop, J.Strang, P.Griffiths, B.Powis and C.Taylor presents an approach in estimating the prevalence of cocaine use, based upon a new ratio estimation technique. This method can be applied to random samples of overlapping populations for which no sampling frames exist. When the ratio estimation method is applied to the two study samples (drawn from populations of people using cocaine and people using heroin) the ratio of cocaine users to heroin users (C/H) was 1.55, with a 95% confidence interval of +/- 0.48. Such estimates should be applied with caution. However, if used with reference to national estimates of about 75,000 heroin users, application of the present estimate suggests that there may be about 116,000 cocaine users in the UK.

Myers and Thompson (1989) pioneered the concept of a generalized approach to estimating hedge ratios, pointing out that the model specification could have a large impact on the hedge ratio estimated. While a huge empirical literature exists on estimating hedge ratios, the literature lacks a formal treatment of model specification uncertainty. These researches accomplish the task by taking a Bayesian approach to hedge ratio estimation, where specification uncertainty is explicitly modeled. Specifically, a Bayesian approach to hedge ratio estimation that integrates over model specification is uncertain; it yields an optimal hedge ratio estimator that is robust to possible model specification because it is an average across a set of hedge ratios conditional on different models. Model specifications vary by exogenous variables (such as exports, stocks, and interest rates) and lag lengths. The methodology is applied to data on corn and soybeans and results showed potential benefits and insights from such approach.

Damoslog and Tomin (2007) considered the Classical Ratio Estimation as a precise and efficient estimator because it has the lowest coefficient of variation and 15% higher relative efficiency than simple random sampling. They also cited that classical ratio estimator has the least estimated bias and is a consistent estimator with its narrow 95 % confidence interval

.

Regression Estimation

Regression estimation aims to find a relationship between a dependent variable and set of independent variable (Kweon and Kocelman, 2004).

In a study conducted by Wei Chen et.Al (2005), linear regression was used to evaluate the parameters with the least squares estimation to estimate an original model for daily shopping with the consideration of individual influence the estimation was verified to be straight forward and efficient.

Myall (2000) noted a known feature of regression estimation, that is, some data will generate negative and weights, which leads to having a download effect on aggregate estimates.

Valliant (2002) used the general regression estimator to construct variance estimators that are approximately model-unbiased in single estimates.

Barrios (1995) explored the used of various small area estimators for various socio-economic indicators. Regression estimation was noted to be reasonable use.

Watson (1937) from Cochran's Citation (1977) illustrated an example of a general situation in which regression estimates are helpful. Watson used a regression on the leaf area and leaf weight to estimate the average area of the leaves on a plant. The procedure was to weight all the leaves of the plant. The area and the weight of each leaf were determined from the small sample of leaves. The sample area was then adjusted by means of the regression on leaf weight means of the regression on leaf weight.

Usha Rani (2001) used the regression model Y=k (LxP) where Y = fruit surface area in Chili (capsicum annuum L.)Fruit, D= diameters of the fruit and k is the regression coefficient. She found the values of k=1.6827.Thus, one can estimate the fruit diameter x 1.6827.

Palaniswamy (1990) reported that the linear regression model can be employed in the prediction of grain yield per plot in rice and consequently for crop estimation.

Lih-Yuan Deng and RAJ S. Chhikara (1990) cited that in a classroom setting students often find the concept of super population and the assumption of a model somewhat artificial when one needs to estimate the mean of a finite population. They introduce the concept of a finite population decomposition based on a regression fit to the population values and then discuss the bias and variance of each estimator from the sampling design viewpoint. It shows by fitting a regression line of y and x to the finite population, that the leading term of the bias of $\hat{y}_R$ is accounted for it on terms of the intercept of the regression line. On the other hand, using quadratic regression fit, describe the leading term of the bias of $\hat{y}_{ir}$ in terms of the coefficient of the quadratic term. Moreover, the sign of intercept in the fitted regression fit would be indicative of an over-or underestimation of $\overline{Y}$ by $\hat{y}_R$ or $\hat{y}_{ir}$,respectively. They hope that this will add to the understanding of the sampling properties of $\hat{y}_R$ and $\hat{y}_{ir}$ without having to assume the super population structure.

Dorfman, J.H and Sanders, D.R (2004) cited that regression and regression-related procedures have become common in survey estimation. It reviews the basic properties of regression estimators, discuss implementation of regression estimation, and investigate variance estimation for regression estimators. The role of models in constructing regression estimators and the use of regression in non-response adjustment are also explored.

A class of ratio and regression type estimators is given such that the estimators are unbiased for random sampling, without replacement, from a finite population. Nonnegative, unbiased estimators of estimator variance are provided for a subclass. Similar results are given for the case of generalized procedures of sampling without replacement. Efficiency is compared with comparable estimation sample selection methods for this case (Rand Corp Santa Monica Calif).

THEORETICAL FRAMEWORK

Simple Random Sampling (SRS)

Simple random sampling is a method of selecting n units out of the N population elements such that every $\binom{N}{n}$ distinct sample has an equal chance of being selected. (Beligan et al., 2004)

$\overline{Y}_n$ is the unbiased estimator of $\overline{Y}_N$ and it is given by:

$$\overline{Y}_n = \frac{\sum_{i=1}^{n} Y_i}{n} \quad \text{(sample mean)} \tag{1.1}$$

and its sample variance is given by:

$$s^2 = \frac{\sum \left(Y_i - \overline{Y}_n\right)^2}{n-1} \tag{1.2}$$

An estimated estimator of the variance of $\overline{Y}_n$ is

$$V\left(\overline{Y}_n\right) = \frac{(N-n)}{N}\left(\frac{s^2}{n}\right) \tag{1.3}$$

where:

$\left(\frac{N-n}{N}\right)$ is the finite population factor(fpcf).The correction factor is used because with small populations, the greater the sampling fraction n/N, the more in formation there is about the population and the smaller the variance to be derived.

The estimated coefficient of variation (C.V) are usually computed using the estimated standard error (S.E)

$$\text{CV}\ (\overline{Y}_n) = \frac{SE\left(\overline{Y}_n\right)}{\overline{Y}_n} \tag{1.4}$$

where:

$$\text{SE}\ (\overline{Y}_n) = \sqrt{\frac{N-n}{Nn}s^2} = \sqrt{V(\overline{Y}_n)} \tag{1.5}$$

Since $\overline{Y}_N = \dfrac{\sum_{i=1}^{n}\overline{Y}_i}{N} = \dfrac{1}{N}\sum_{i=1}^{N}Y_i$ and to estimate population mean is given by

$$\overline{Y}_N = \text{N}\ \overline{Y}_n \tag{1.6}$$

The variance, standard error (SE) and the coefficient of variation (C.V) of the population total are

$$\hat{V}\ (\text{N}\ \overline{Y}_n) = \frac{N(N-n)}{n}s^2$$

proof:

$$\hat{V}\ (\text{N}\ \overline{Y}_n) = \text{N}^2(\text{V}(\overline{Y}_n))$$

$$= \text{N}^2\left(\frac{N-n}{N}\right)\left(\frac{S}{n}\right)$$

$$= \frac{N\left(N-n\right)}{n}s^2$$

If S^2 is not available, therefore the estimates of $\text{V}(\text{N}\ \overline{Y}_n)$ is

$$\hat{V}(\text{N } \overline{Y}_n) = \frac{N(N-n)}{n}s^2 \tag{1.7}$$

$$\text{SE (N } \overline{Y}_n) = \sqrt{\frac{N(N-n)}{n}s^2} = \sqrt{V(N\overline{Y}_n)} \tag{1.8}$$

Confidence interval is a range of possible values for the unknown parameter with some measure for the degree of certainty, (1-α) called the level of confidence coefficient. Hence, the confidence interval for the mean ($\overline{Y}_n$) and total (N$\overline{Y}_n$) are as follows:

Confidence Interval of the mean estimate:

$$P\left[\overline{Y}_n - t_{\frac{\alpha}{2},n-1} SE\left(\overline{Y}_n\right) \leq \overline{Y}_n \leq \overline{Y}_n + t_{\frac{\alpha}{2},n-1} SE\left(\overline{Y}_n\right)\right] = 1-\alpha \tag{1.9}$$

Confidence interval of the total estimate:

$$P\left[N\overline{Y}_n - t_{\frac{\alpha}{2},n-1} SE\left(N\overline{Y}_n\right) \leq N\overline{Y}_n \leq N\overline{Y}_n + t_{\frac{\alpha}{2},n-1} SE\left(N\overline{Y}_n\right)\right] = 1-\alpha \tag{1.1}$$

<u>Classical Ratio Estimation</u>

In ratio estimation, the ratio estimate of $\overline{Y}_N$ using SRS is

$$\overline{Y}_R = \frac{\overline{Y}_n}{\overline{X}_n}\overline{X}_N \tag{2.1}$$

where

$\overline{X}_N$ (Total area of the strawberry)

$$\overline{Y}_n = \frac{\sum_{i=1}^{n} y_i}{n} \text{ (Sample mean for the production of strawberry)} \tag{2.2}$$

$$\overline{X}_n = \frac{\sum_{i=1}^{n} x_i}{n} \text{ (Sample mean of the number of strawberry plant)} \qquad (2.3)$$

According to Hartley and Ross, $\overline{R}_n$ is biased for $\overline{R}_N$,and the bias is

$$\text{bias } (\overline{R}_n) = \left(\frac{N-n}{Nn\overline{x}_n^2}\right) \left\{\overline{R}_n s_x^2 - s_{xy}\right\} \qquad (2.4)$$

where:

$$\overline{R}_n = \frac{\overline{Y}_n}{\overline{X}_n} \quad \text{(sample ratio of the mean)} \qquad (2.5)$$

$$s_x^2 = \frac{\sum (x_i - \overline{x}_n)^2}{n-1} \quad \text{(sample variance of } x_i \text{)} \qquad (2.6)$$

$$s_{xy}^2 = \frac{\sum \left(x_i - \overline{x}_n\right)\left(y_i - \overline{y}_n\right)}{n-1} \quad \text{(sample covariance of } x_i \text{ and } y_i \text{)} \qquad (2.7)$$

Hence,

$$\text{Bias } (\overline{Y}_R) = \overline{X}_N \text{ bias } (\overline{R}_n) \qquad (2.8)$$

The variance of classical ratio estimate and its standard error are

$$V(\overline{Y}_R) = \frac{N-n}{Nn(n-1)} \sum \left[y_i - \overline{R}_n x_i\right]^2 \qquad (2.9)$$

$$SE(\overline{Y}_R) = \sqrt{V\left(\overline{Y}_R\right)} \qquad (2.10)$$

Estimate coefficient of variation of the estimate is estimated using the S.E. but the square root of the MSE will be used to estimate the CV where the variance is adjusted for its bias:

$$\mathrm{CV}(\bar{Y}_R) = \sqrt{\frac{MSE(\bar{Y}_R)}{Y_R}}100 \qquad (2.11)$$

where:

$$\mathrm{MSE}\ (\bar{Y}_R) = \mathrm{V}\ (\bar{Y}_{\mathrm{R}}) + \mathrm{bias}\ ((\bar{Y}_R))^{2} \qquad (2.12)$$

To estimate the population mean

$$\bar{Y}_N = \mathrm{N}\bar{Y}_{\mathrm{R}} \text{ is used} \qquad (2.13)$$

The variance, standard error (S.E.) and bias of $\mathrm{N}\bar{Y}_{\mathrm{R}}$ are as follows:

$$\mathrm{V}\ (\mathrm{N}\bar{Y}_{\mathrm{R}}) = \mathrm{N}^2\ \mathrm{V}\ (\bar{Y}_{\mathrm{R}}) \qquad (2.14)$$

$$\mathrm{SE}\ (\mathrm{N}\bar{Y}_R) = \sqrt{V(N\bar{Y}_R)} \qquad (2.15)$$

and bias of population total is given by

$$\mathrm{bias}\ (\mathrm{N}\bar{Y}_R) = \mathrm{N}(\mathrm{bias}\ (\bar{Y}_{\mathrm{R}})) = \mathrm{N}[\ \bar{X}_N\ \mathrm{bias}\ (\bar{R}_n)\ ] \qquad (2.16)$$

The confidence interval for the mean and total are shown below:

Confidence Interval of the mean estimate:

$$P\left[\bar{Y}_n - t_{\frac{\alpha}{2},n-1} SE(\bar{Y}_n) \le \bar{Y}_n \le \bar{Y}_n + t_{\frac{\alpha}{2},n-1} SE(\bar{Y}_n)\right] = 1-\alpha \qquad (2.17)$$

Confidence Interval of the total estimate:

$$P\left[N\bar{Y}_n - t_{\frac{\alpha}{2},n-1} SE(N\bar{Y}_n) \le N\bar{Y}_n \le N\bar{Y}_n + t_{\frac{\alpha}{2},n-1} SE(N\bar{Y}_n)\right] = 1-\alpha \qquad (2.18)$$

Hartley-Ross Ratio-Type Estimation

Instead of using $\overline{R}_n = \dfrac{\overline{Y}_n}{\overline{X}_n}$,we use

$$\overline{r_n} = \frac{1}{n}\sum_{i=1}^{n} r\underline{i} \qquad \text{(sample mean of ratio)} \tag{3.1}$$

Where:

$$r_i = \frac{y_i}{x_i} \text{;(individual sample ratio)} \tag{3.2}$$

$$r_1 = \frac{y_2}{x2}$$

$$r_2 = \frac{y_2}{x_2}$$

$$\vdots$$

$$r_n = \frac{y_n}{x_n}$$

Note that E($\overline{r}_n$) $\neq \overline{R}_n \Rightarrow \overline{r}_n$ is inconsistent, therefore $\overline{Y}_{\mathrm{r}} = \overline{r}_n \overline{X}_N \Rightarrow \overline{Y}_{\mathrm{r}}$ is biased and consistent. Bias estimator of $\overline{Y}_{\mathrm{r}}$ is given as

$$\text{Bias}(\overline{Y}_{\mathrm{r}}) = \frac{N-1}{N}(s_{rx}) \tag{3.3}$$

Proof:

$$\text{Bias}(\overline{Y}_{\mathrm{r}}) = (\overline{X}_N \overline{r}_N) - \overline{Y}_N = (\overline{X}_N \overline{r}_N) - \frac{\sum y_i}{N}$$

$$= \frac{1}{N}\left[\sum y_i - N\overline{X}_N \overline{r}_N\right]$$

$$= \frac{1}{N}\left[\sum r_i x_i - N\overline{X}_N \overline{r}_N\right]$$

$$= \frac{1}{N}\left[\sum r_i x_i - N\frac{\sum X_i}{N}\frac{\sum r_i}{N}\right]$$

$$= \frac{1}{N}\left[\sum r_i x_i - N\frac{\sum X_i \sum r_i}{N}\right]$$

Multiplying by N-1

$$= -\ \frac{N-1}{N}\left[\sum r_i x_i - \frac{\sum X_i \sum r_i}{N}\right]$$

$$= -\ \frac{N-1}{N}(S_{rx})$$

$$\text{Bias}(\bar{Y}_{\text{r}}) = -\ \frac{N-1}{N}(s_{rx})$$

Wherein the unbiased estimator of s_{rx}

$$S_{\text{rx}} = \frac{1}{n-1}\left[\bar{y}_n - \bar{r}_n \bar{x}_n\right] \tag{3.4}$$

Therefore,

$$\text{Bias}(\bar{Y}_{\text{r}}) = -\ \frac{N-1}{N}\left(\bar{y}_n - \bar{r}_n \bar{x}_n\right) \tag{3.5}$$

To get the unbiased estimator which make use of $\bar{r}_n$,subtract the bias from $\bar{y}_r = \bar{r}x_N$, therefore,

$$\bar{y}_r = \bar{r}\bar{X}_N + \frac{(N-1)n}{N(n-1)}\left(\bar{y}_n - \bar{r}_n \bar{x}_n\right) \tag{3.6}$$

As cited by Hartley and Ross, the formula of the variances of the ratio type estimation for large sample is given by

$$\mathrm{V}(\bar{y}_r) = \frac{1}{n}\frac{(N-1)}{N}\left(S_y^2 - 2\bar{r}_N S_x^2\right) + \frac{1}{n(n-1)}\frac{(N-1)^2}{N^2}\left(S_{rx}^2 + S_r^2 S_x^2\right)$$

The unbiased estimator is given by

$$\hat{V}\ (\bar{y}_r) = \frac{1}{n}\frac{(N-1)}{N}\left(S_y^2 - 2\bar{r}_N S_x^2\right) + \frac{1}{n(n-1)}\frac{(N-1)^2}{N^2}\left(S_{rx}^2 + S_r^2 S_x^2\right) \qquad (3.7)$$

Where

$$S_x^2 = \frac{\sum\left(x_i - \bar{x}_n\right)^2}{n-1} \qquad (3.8)$$

$$S_y^2 = \frac{\sum\left(y_i - \bar{y}_n\right)^2}{n-1} \qquad (3.9)$$

$$S_r^2 = \frac{\sum\left(r_i - \bar{r}_n\right)^2}{n-1} \qquad (3.10)$$

$$S_{rx} = \frac{n}{n-1}\left[\bar{y}_n - \bar{r}_n \bar{x}_n\right] \qquad (3.11)$$

$$S_{xy}^2 = \frac{\sum\left(x_i - \bar{x}_n\right)\left(y_i - \bar{y}_n\right)}{n-1} \qquad (3.12)$$

Estimated standard error is,

$$SE(\bar{y}_r) = \sqrt{V\left(\bar{y}_r\right)} \qquad (3.13)$$

This is the square root of the variance. The estimated coefficient of variation of the estimate is shown below,

$$CV(\bar{y}_r) = \sqrt{\frac{MSE(\bar{y}_r)}{\bar{y}_r}}100 \tag{3.14}$$

Where,

$$MSE(\bar{y}_r) = V(\bar{y}_r) + (Bias(\bar{y}_r))^2 \tag{3.15}$$

All these result apply to the estimation of the population total $\bar{Y}_N$.The unbiased estimator of $\bar{Y}_N$ = N$\bar{Y}n$ and the variance of standard error(SE) and bias of N$\bar{Y}n$ is defined by

$$V(N\bar{Y}_r = N^2 V(\bar{Y}_r) \tag{3.16}$$

$$SE(N\bar{Y}_r) = \sqrt{V(N\bar{y}_r)} \tag{3.17}$$

$$Bias(N\bar{Y}_r) = N(Bias(\bar{Y}_r)) = -N\frac{N-1}{N}(s_{rx}) \tag{3.18}$$

The confidence interval for the mean and total are shown below confidence interval for the mean ($\bar{Y}_r$)

Confidence Interval of the Mean

$$P\left[\bar{Y}_r - t_{\frac{\alpha}{2},n-1} SE(\bar{Y}_r) \leq \bar{Y}_r \leq \bar{Y}_r + t_{\frac{\alpha}{2},n-1} SE(\bar{Y}_r)\right] = 1-\alpha \tag{3.19}$$

Confidence Interval of the Total

$$P\left[N\bar{Y}_r - t_{\frac{\alpha}{2},n-1} SE(N\bar{Y}_r) \leq N\bar{Y}_r \leq N\bar{Y}_r + t_{\frac{\alpha}{2},n-1} SE(N\bar{Y}_r)\right] = 1-\alpha \tag{3.20}$$

<u>Regression Estimation</u>

The linear regression estimator is designed to increase precision by the use of auxiliary variant x_i that is correlated with y_i. Suppose y_i and x_i are obtained for every unit in the sample that the population mean $\overline{X}_N$ of x_i is known then the regression estimator of $\overline{Y}_N$ is

$$\overline{Y}_{reg} = \overline{y}_n + \beta\left(\overline{X}_N - \overline{x}_n\right) \tag{4.1}$$

Where the subscript reg denote as regression and β is the estimate of the change in y and x when x is increased by unity. The model is shown,

$$\beta = \frac{\sum\left(x_i - \overline{x}_n\right)\left(y_i - \overline{y}_n\right)}{\sum_{i=1}^{n}\left(x_i - x_n\right)^2} \tag{4.2}$$

Where,

$$\overline{Y}_n = \frac{\sum_{i=1}^{n}\overline{Y}_i}{n} \quad \text{(Sample mean for the production of strawberry)} \tag{4.3}$$

$$\overline{X}_n = \frac{\sum_{i=1}^{n}x_i}{n} \quad \text{(Sample mean of the number of strawberry plant)} \tag{4.4}$$

Variance of regression estimation is estimated using the variance of residual as

$$V(\overline{Y}_{reg}) = \left(1 - \frac{n}{N}\right)\frac{se^2}{n} \tag{4.5}$$

Where

$$e_i = y_i - \bar{Y}_n + \hat{\beta}_1\left(x_i - \bar{x}_n\right) \text{ (residual sample)} \tag{4.6}$$

where,

$$\hat{\beta}_1 = \frac{n\sum x_i - y_i - \sum\left(x_i\right)\left(y_i\right)}{n\sum {}_i^2 - \left(\sum x_i\right)^2} \text{ (slope)} \tag{4.7}$$

$$\bar{e}_n = \frac{1}{n}\sum e_i \text{ (Mean of the residual)} \tag{4.8}$$

$$s_e^2 = \frac{\sum\left(e_i - \bar{e}_n\right)^2}{n-1} \text{ (Variance of the residual)} \tag{4.9}$$

Standard error is defined by

$$SE\left(\bar{Y}_{reg}\right) = \sqrt{V\left(\bar{Y}_{reg}\right)} \tag{4.10}$$

And the bias of $\left(\bar{Y}_{reg}\right)$ is estimated as

$$\text{bias}\left(\bar{Y}_{reg}\right) = \frac{1 - \frac{n}{N}}{n_{sx}^2}\sum_{i=1}^{N}\frac{e_i\left(x_i - \bar{x}_n\right)^2}{N-1} \tag{4.11}$$

The coefficient of variation of the estimator $\left(\bar{Y}_{reg}\right)$ which measures the variability of the estimate is,

$$CV\left(\bar{Y}_{reg}\right) = \frac{\sqrt{MSE}\left(\bar{Y}_{reg}\right)}{\bar{Y}_{reg}}100 \tag{4.12}$$

where

$$MSE\left(\overline{Y}_{reg}\right) = V\left(\overline{Y}_{reg}\right) + \left(Bias\left(\overline{Y}_{reg}\right)\right)^2 \qquad (4.13)$$

To estimate the population mean, $\overline{Y}_N$, we used the unbiased estimator

$$\overline{Y}_N = \mathrm{N}\left(\overline{Y}_{reg}\right) \qquad (4.14)$$

And estimate variance of the total $\mathrm{N}\left(\overline{Y}_{reg}\right)$ is

$$\mathrm{V}\left(N\overline{Y}_{reg}\right) = \mathrm{N}^2\,\mathrm{V}\left(\overline{Y}_{reg}\right) \qquad (4.15)$$

$$SE\left(N\overline{Y}_{reg}\right) = \sqrt{V\left(N\overline{Y}_{reg}\right)} \qquad (4.16)$$

And its bias is given by

$$bias\left(N\overline{Y}_{reg}\right) = \mathrm{N}\left(bias(\overline{Y}_{reg})\right) = -\mathrm{N}\ \frac{1-\frac{n}{N}}{n_{sx}^2}\sum_{i=1}^{N}\frac{e_i\left(x_i - \overline{x}_n\right)^2}{N-1} \qquad (4.17)$$

At the regression estimation in the measure of accuracy of the estimate is approximately 95% confidence interval(C.I) where half width of the C.I is the margin error of an estimate that is,

Confidence Interval of the mean:

$$P\left[\overline{Y}_{reg} - t_{\frac{\alpha}{2,n-1}}SE\left(\overline{Y}_{reg}\right) \le \overline{Y}_{reg} \le \overline{Y}_{reg} + t_{\frac{\alpha}{2,n-1}}SE\left(\overline{Y}_{reg}\right)\right] = 1-\alpha \qquad (4.18)$$

Confidence Interval of the total

$$P\left[N\overline{Y}_{reg} - t_{\frac{\alpha}{2},n-1} SE\left(N\overline{Y}_{reg}\right) \leq N\overline{Y}_{reg} \leq N\overline{Y}_{reg} + t_{\frac{\alpha}{2},n-1} SE\left(N\overline{Y}_{reg}\right) \right] = 1 - \alpha$$

<u>Magnitude of Efficiency</u>

This is used to determine the efficiency of each estimator with respect to the y-only estimator. The magnitude of efficiency can be computed using the formula bellow:

RE(Classical ratio over "Y-only", in %) = $\frac{V\left(\overline{Y}_n\right)}{V\left(\overline{Y}_R\right)}(100)$ (5.1)

RE (Hartley-Ross over "Y-only", in %) = $\frac{V\left(\overline{Y}_n\right)}{V\left(\overline{Y}_r\right)}(100)$ (5.2)

RE (Regression over "Y-only", in %) = $\frac{V\left(\overline{Y}_n\right)}{V\left(\overline{Y}_{reg}\right)}(100)$ (5.3)

Definition of Terms

Strawberry. Pulpy red fruit with a seeded-studded surface plant with runners and white flower bearing this.

Production. Something produced.

Estimate. Approximate judgment, especially of cost, value, size, ect.statement of approximate charge for work to be undertaken

Estimation. The process of providing a numerical value for population parameter based on information collected from a sample.

Estimator. A statistic used to provide an estimate for a parameter. A sample mean for example is of the population mean.

Ratio Estimation. The use of the ratio estimator when the relationship between the response y and the subsidiary variable x which is proportional to y.

Regression Estimation: It is the relationship between the mean value of a random variable and the corresponding values of one or more independent variables.

Variable .Used to estimate the relationship between the y's and the x's not through the origin but is a straight line.

Subsidiary variable. Used to estimate the mean of the response y.

Dependent variable. It is the variable to be determined or explained by one or more explanatory variable.

METHODOLOGY

Sampling Design

Simple random sampling was employed in the selection of sample plots using random numbers. The population consisted of 899 plots from 11,488 square meters of the Sariling Sikap area and 22 farmers. The plots were planted with strawberries.195 plots were drawn as samples. Each plot has different number of strawberry plants and sizes of plots.

Data Gathering

The data was gathered through harvesting strawberries. The researchers asked permission from the farmers to harvest the strawberry from the selected plots and weighted it. The researchers gathered data until the sixth harvesting season of the strawberry. Data gathering started on the 29th day of December, 2007 and ended on the 3rd day of February, 2008.Harvest is done twice a week since farmers must spray and exposed the plants for two days before the harvesting.

Data Analysis

The data gathered were summarized using the Microsoft excel program.

The mean, variance, mean square error, and bias of ratio, regression and simple random sampling estimates were computed and compared from each other through their relative efficiency with the strawberry production and number of plants as auxiliary variable used.

RESULTS AND DISCUSSION

The Estimates of Strawberry Production/grams
in Strawberry Farm, La Trinidad, Benguet

The estimates of the mean and total production of strawberry using the four estimators are presented in Table 1. It showed that simple random sampling had the least estimated mean that is about 7,208.718 grams and a total of 6,480,637.482 grams in estimating strawberry production. Hartley-Ross estimation had the largest mean and total estimation that is about 9,888.170 grams and 8,889,464.170 grams. Classical estimation had 9,844.239 grams estimated mean and 8,849,970.861 grams estimated total production while regression estimation had the largest mean and total estimation that is about 9,888.170 grams and 8,889,464.170 grams. Classical estimation had 9,844.239 grams estimated mean and 8,849,970.861 grams estimated total production while regression estimation has 8,613.167 grams and 7, 743.133 grams estimated mean and total production of strawberry per plot in Strawberry Farm.

According to Cochran (1977) simple random sampling is a method of selecting n units out of the N such that every one of the ${}_NC_n$ distinct samples has an equal chance of being drawn. He stated also that Ratio estimate is consistent. It is biased except for some special types of population, although the bias is negligible for large samples.

Table1. Estimates of mean and total of Strawberry Production using four methods of estimation.

ESTIMATION PROCEDURE	ESTIMATOR	ESTIMATE	±	STANDARD ERROR	BIAS
for average					
SRS	$\overline{Y}_n$	7,208.72	±	209.331	0
Classical Ratio	$\overline{Y}_{\mathrm{R}}$	9,844.239	±	207.501	7.218
Hartley-Ross	$\overline{Y}_{\mathrm{r}}$	9,888.170	±	239.867	121.99
Regression	$\overline{Y_{ir}}$	8,613.167	±	201.426	0.0223
for total					
SRS	$\mathrm{N}\overline{Y}_n$	6,480,637.482	±	188,188.469	0
Classical Ratio	$\mathrm{N}\overline{Y}_{\mathrm{R}}$	8,849,970.861	±	186,548.308	6,488.982
Hartley-Ross	$\mathrm{N}\overline{Y}_{\mathrm{r}}$	8,889,464.170	±	516,943.809	109,669.01
Regression	$\mathrm{N}\overline{Y_{ir}}$	7,743,237.133	±	181,082.317	20.048

Consistency of the Estimates

Table 2 shows the Confidence Interval of the true mean and total of the strawberry production using the different estimation procedures. Result shows that Hartley-Ross ratio-type has the largest 95% Confidence Interval for the true mean with a value of 9,598.167 grams to 10,357.957 grams and true total with a value indicated to be 8,467,126.872 grams to 9,311,802.788. Simple random sampling, a biased estimator has the least confidence interval for true mean with a values of 6,798.429 to 7,619.007 grams and total production of 6,111,788.083 grams to 6,849,486.881 grams. Moreover, for the bias estimators, regression

Table 2. Confidence Interval of the true mean and total strawberry production using three methods of estimation

ESTIMATION PROCEDURE	LOWER LIMIT	UPPER LIMIT
for average		
SRS	6,798.429	7,619.007
Classical Ratio	9,473.537	10,250.941
Hartley-Ross	9,598.167	10,421.065
Regression	8,218.372	9,007.962
for total		
SRS	6,111,788.083	6,849,486.881
Classical Ratio	8,484,345.977	9,215,595.743
Hartley-Ross	8,628,750.649	10,011,902.3
Regression	7,388,315.792	8,098,158.474

estimator has the least confidence interval with 8,218.372 to 9,007.962 for the true mean and 7,388,315.880 to 8,098,158.386 for the true total production.

Precision of the Estimates

The study of Damoslog and Tomin (2007) showed that classical ratio estimator with the least variability is the more precise estimator. Table 3 showed that classical ratio estimation had the least variability with 2.108 is the more precise estimation but regression estimator also had a close value of coefficient of variation which is 2.339. Specifically, simple random sampling has the highest coefficient of variation with a value of 2.904 compared to the other estimators. It was also found that the different estimators have small variability with Hartley

Ross ranked second from the highest variability obtained followed by regression with 2.339 coefficient of variation.

Table3. Coefficient of Variability of the four methods of estimation.

ESTIMATION PROCEDURE	COEFFICIENT OF VARIATION
SRS	2.904
Classical Ratio	2.108
Hartley-Ross	2.720
Regression	2.339

Efficiency of the Estimates

This is supported by Cochran (1977) as cited by Beligan (2004). If $\hat{o}_1$ and $\hat{o}_2$ are unbiased estimators for the parameters θ, that is E($\hat{o}_1$) = E($\hat{o}_2$) = θ, then $\hat{o}_1$ is more efficient than $\hat{o}_2$ if V($\hat{o}_1$) < V($\hat{o}_2$).

In the study of Damoslog and Tomin (2007), an estimator with higher relative efficiency is considered an efficient Thus, in the said study classical ratio was efficient compared to simple random sampling.

Based from the mentioned criterion for efficient estimators, regression is the most efficient estimators in estimating strawberry production with a variance of 40,572.592 and higher relative efficiency as shown in Table 4. Hartley-Ross ratio-type having a value of 57,449.827 has the largest variance in estimating strawberry production and classical ratio is next to the regression estimation with a variance of 43,189.421.

And this result is the same with the study of Barrios (1995) where regression estimator was considered efficient in estimating socio-economic indicators.

Table 4. Relative Efficiency and Variance of the different estimates compared to simple random sampling.

PROCEDURE	RELATIVE EFFICIENCY	VARIANCE
SRS		
Classical Ratio	101.77	43,189.421
Hartley-Ross	76.274	57,449.827
Regression	108	40,572.592

SUMMARY, CONCLUSION AND RECOMMENDATION

Summary

The study was conducted to determine the efficiency and consistency of ratio and regression estimation in estimating the strawberry production in Strawberry farm at La Trinidad, Benguet over the simple random sampling method.

The data was gathered from the 29th day of December 2007 until 3rd day of February 2008 wherein the researchers ask permission from the owner of the strawberry that was selected as one of the study site. The sample plots were determined using simple random sampling and there were about 195 plots considered. The data gathering ended until the sixth harvesting of strawberries.

The production was estimated using simple random sampling classical ratio estimation, Hartley-Ross ratio-type and regression estimation. It was found that each estimated values obtained due to different estimation schemes utilized.

The classical ratio, Hartley–Ross ratio-type and regression estimators have greater estimated mean and total than the simple random sampling. The regression estimator was said to have the least bias while the classical ratio was considered to have the most bias. Though simple random sampling estimator is not bias, it is considered as a consistent estimator since its 95 % confidence interval for the true mean and true total of the strawberry production is narrow. Among the considered bias estimator, regression was more consistent with its 95 % confidence interval which is limited or having the least confidence interval. Classical ratio estimate is also considered as the most precise among the estimates since it has the lowest coefficient of variation. Result of the relative efficiency shows that regression estimator is the more efficient compared to simple random sampling.

Conclusion

It is therefore concluded that classical ratio estimator is the most precise among the four different estimators used in estimating strawberry production with its least coefficient of variability. Moreover, simple random sampling was a consistent and unbiased estimator.

However, regression estimator is the most efficient and most consistent among the bias estimators as applied to strawberry production.

Recommendation

It is recommended that some study has to be conducted using the independent variable that is the number of plants per plots to determine the efficiency of estimator for more meaningful estimate .To maintain a high degree of precision, the variability among sampling units within a plot should be kept small. It is also recommended that in estimating strawberry production it is efficient to use the regression estimator over simple random sampling.

LITERATURE CITED

BARRIOS, E. B. 1998. Small area Estimator of Employment Rate.7th National Convention on Statistics. Manila, Philippines.

BELIGAN, S.Z. 2004. Survey and Sampling Techniques Syllabus

COCHRAN, G. W. 1997.Sampling Techniques.3rd Edition. John Wiley and Sons, United States of America.

DAMOSLOG, F.T. AND TOMIN, J.T. 2007. Efficiency of Ratio and Regression Estimator s in Estimating Farmers' Income in La Trinidad Benguet.

DENG, LI-YUAN AND CHHIKARA.R.S (1990). On the Ratio and Regression Estimation in Finite Population Sampling

DORFMAN, J.H. AND SANDERS, D.R. 2004. Generalized Hedge Ratio Estimation with an unknown model

EVERITT, B.S. 1998. Dictionary of Statistics.

FAEDI W., MOURGUES F AND ROSATI. Strawberry Breeding and Varieties: Situation and Perspective

HSU AND KUO. 2000. Household Solid Waste Recycling Included Production.

LOHR,S.L.1999. Sampling:Design and Analysis.Brooks/Cole PublishingCompany

M.GOSSOP, J.STRANG, P.GRIFFITHS, B.POWIS AND C.TAYLOR. Drug Transitions Study, National Addiction Centre, Maudsley Hospital, London

ONATE, B.T.1967. Ratio Estimators in Field Research. The Philippine Statistician

Vol.VXI, No.3-4:62-74

(Oxford Quick Reference)Dictionary

PALANISWAMY, U.R AND PALANISWAMY, K.M.(2006). Handbook of Statistics for teaching and research in plant and crop sciences.

RAND CORP SANTA MONICA CALIF. Some Finite Population Unbiased Ratio and Regression Estimators.

SUKHATME, B.V. Hartley-Ross Unbiased Ratio-Type Estimator

TOTEMIC N. All season Strawberry Growing with day-neutral Cultivars.

WALPOLE, R.E.ET.AL.1998. Probability and Statistics for Engineers and Scientists.Prentice-Hall, Inc.6th edition.

WEI CHEN, YOSHINAO OOEDA AND TOMONORI SUMI. 2005. A study on a shopping Frequency Model with the Consideration of Individual Difference. Kyushu University, Japan.

APPENDICES

Appendix A. Application for Oral Defense

Benguet State University
College of Arts and Sciences
DEPARTMENT OF MATHEMATICS-PHYSICS-STATISTICS
APPLICATION FOR ORAL DEFENSE

Date: March 17, 2008

Group Members:

Emily June C. Alos — Major: Statistics
Eleonor B. Bat-oy — Minor: Information Technology

Degree: BACHELOR OF SCIENCE AND APPLIED STATISTICS
Title of Thesis: "Efficiency of Ratio and Regression Estimation in Estimating the Strawberry Production."

ENDORSED BY : MARYCEL H. TOYHACAO
Adviser

Date and time of Defense: March 18, 2008 @ 5 PM
Place of Defense: CAS Annex 210

NOTED BY : MARIA AZUCENA B.LUBRICA
Department Chairman

REPORT OF RESULT ON ORAL DEFENSE

Name and signature	*Remarks
Marycel H. Toyhacao Adviser	________________________
Salvacion Z. Beligan Member	________________________
Maria Azucena B.Lubrica Member	________________________
Cristina B.Ocden Member	________________________

*passed or failed

Appendix B. The raw data of strawberry production

				Harvesting				
plots	1st	2nd	3rd	4th	5th	6th	Production(g)	# of plants
1	1000	700	700	1000	1700	1100	6200	120
2	1000	600	800	500	1700	1100	5700	116
3	1000	900	500	600	1100	1200	5300	114
4	1000	500	1300	900	1900	1800	7400	122
5	700	700	800	500	900	1000	4600	110
6	800	700	900	600	1600	1250	5850	110
7	1000	500	1100	900	1700	1600	6800	122
8	1000	300	1000	1000	1300	1400	6000	118
9	1000	600	1100	1500	1200	1100	6500	116
10	700	500	600	1500	1200	1500	6000	116
11	700	400	700	1000	1100	1600	5500	118
12	700	600	800	1200	1300	1300	5900	120
13	600	400	1000	1000	1300	1250	5550	136
14	600	600	600	500	1400	1100	4800	110
15	1200	600	600	900	1200	1600	6100	112
16	600	600	700	500	1400	1750	5550	112
17	600	600	600	1100	1250	1900	6050	101
18	500	500	900	1200	1700	1100	5900	112
19	700	200	900	1500	2300	1150	6750	118
20	500	600	800	800	1700	1900	6300	120
21	600	400	800	700	1300	1800	5600	118
22	700	400	900	900	1400	1400	5700	112
23	600	400	900	1700	1450	1450	6500	110
24	700	400	900	1700	1950	1000	6650	108
25	600	300	800	1400	1850	1100	6050	110
26	600	500	900	1800	1500	1150	6450	112
27	500	300	900	1400	1900	1250	6250	106
28	600	500	600	1700	1500	1000	5900	130
29	700	500	900	1700	1850	900	6550	114
30	600	600	600	1600	1700	1800	6900	112
31	500	500	800	1600	1700	1800	6900	114
32	500	500	500	1600	1450	1900	6450	110
33	700	300	1200	1300	1700	1100	6300	114
34	500	300	700	1350	1450	1200	5500	106
35	600	300	800	1250	1500	1000	5450	110
36	400	300	700	1000	1700	1700	5800	130
37	400	500	800	1600	1200	1600	6100	108
38	900	500	1100	600	2250	1850	7200	110
39	600	400	1000	1400	1900	1550	6850	104
40	300	800	900	600	1300	1250	5150	106
41	500	200	700	600	900	1900	4800	124
42	300	100	400	800	1100	1800	4500	106

43	600	400	800	400	300	1000	3500	118
44	400	600	600	400	500	1150	3650	120
45	400	500	900	500	500	1100	3900	106
46	500	400	400	300	500	1000	3100	108
47	600	300	700	300	400	1500	3800	106
48	700	800	100	400	400	1000	3400	108
49	300	600	1000	400	500	800	3600	106
50	500	300	1200	500	500	1100	4100	94
51	400	400	700	400	300	1250	3450	108
52	300	500	900	300	300	1200	3500	104
53	400	700	900	500	500	800	3800	108
54	700	200	600	200	300	900	2900	110
55	500	600	500	200	300	1150	3250	106
56	400	400	700	300	400	800	3000	108
57	400	400	600	400	400	850	3050	118
58	400	600	700	200	400	1050	3350	116
59	600	700	600	300	300	850	3350	110
60	500	300	700	400	500	700	3100	108
61	800	900	800	500	500	1350	4850	108
62	400	400	600	400	400	800	3000	106
63	400	200	400	400	500	1250	3150	110
64	400	300	400	200	300	1000	2600	112
65	200	200	300	200	200	1000	2100	120
66	400	500	300	300	300	1100	2900	130
67	150	200	200	300	400	1150	2400	118
68	250	250	300	100	200	800	1900	114
69	200	300	400	200	200	1150	2450	116
70	400	300	200	100	300	1000	2300	122
71	200	300	400	300	400	1250	2850	120
72	500	300	300	300	500	800	2700	120
73	400	300	300	300	400	800	2500	124
74	300	400	500	200	100	900	2400	126
75	200	500	200	400	300	950	2550	130
76	300	500	200	100	100	850	2050	126
77	300	300	200	300	400	800	2300	128
78	200	400	500	400	300	1000	2800	129
79	300	300	200	200	400	1100	2500	132
80	600	500	700	700	500	1150	4150	130
81	500	900	900	900	900	800	4900	128
82	300	900	900	900	800	700	4500	122
83	1300	17000	900	1100	1200	2000	23500	115
84	1100	8000	900	1400	1400	2100	14900	96
85	1600	1000	500	1000	1600	2000	7700	116
86	1200	900	300	1000	1000	1500	5900	114
87	11100	1000	1200	1500	1900	2200	18900	123
88	1200	1000	600	1400	1600	2000	7800	120

89	1700	11000	1100	1500	1800	2300	19400	115
90	2200	1500	1700	2100	1900	2350	11750	117
91	1200	1000	1000	1600	2100	2500	9400	116
92	1700	1500	1200	1350	2100	2100	9950	107
93	1600	1500	1600	1700	1800	1500	9700	103
94	800	1300	1600	1900	1900	2200	9700	112
95	11000	1500	1500	2000	1300	2000	19300	105
96	1100	1500	800	1500	1600	1800	8300	106
97	1100	1100	500	1600	1800	1800	7900	112
98	1300	900	800	1400	1700	2000	8100	120
99	1000	900	1000	1500	1600	1800	7800	93
100	1400	900	900	1800	1700	2100	8800	112
101	12000	1200	1000	1900	2000	1200	19300	117
102	1100	1000	900	2200	2300	1250	8750	110
103	1300	1200	1000	1300	1500	2000	8300	119
104	1100	1300	1100	1600	1800	1400	8300	115
105	1400	1200	1000	1200	1600	1400	7800	114
106	1700	1800	1500	1400	1500	1450	9350	117
107	1200	1300	1500	1500	1300	1000	7800	115
108	1600	1200	1100	1600	1800	1100	8400	114
109	1300	1000	900	1800	1900	1000	7900	117
110	1500	1500	1300	1900	1600	1600	9400	115
111	1000	1500	1400	1800	1700	1900	9300	115
112	1500	1000	800	1300	1400	2000	8000	120
113	1300	1300	1100	1400	1600	2200	8900	120
114	1500	1600	1500	1700	1000	2250	9550	113
115	1600	1500	1200	1600	2000	2000	9900	118
116	1800	1500	1500	1800	2100	2200	10900	120
117	2200	2000	1800	2200	2200	2400	12800	134
118	1300	1500	1100	2100	2300	2000	10300	124
119	1200	1100	1000	1400	2100	1500	8300	123
120	1300	1400	1300	1500	1900	1600	9000	116
121	1000	1000	800	1600	1700	1800	7900	118
122	1000	1200	1500	1800	2000	1600	9100	101
123	1500	1400	1300	2100	2100	2000	10400	121
124	1500	1600	1500	2400	2100	1800	10900	112
125	1400	1200	1300	1400	1500	2200	9000	124
126	1500	1300	1000	1600	1900	2300	9600	105
127	1000	1000	800	1800	2000	2000	8600	100
128	1300	1300	1000	1000	1200	2250	8050	102
129	1400	2500	2400	2600	2300	2000	13200	150
130	1300	1100	1100	1300	1400	1250	7450	145
131	1600	1000	1600	1700	1400	1400	8700	142
132	1200	1100	1500	1800	1900	1800	9300	132
133	1300	1300	1600	1400	1500	1200	8300	131
134	1600	1200	1400	1200	1100	1000	7500	142

135	1700	700	1900	2000	1800	1500	9600	144
136	1400	700	1500	1300	1400	1900	8200	145
137	1200	750	1600	1350	1400	1400	7700	138
138	1300	1000	1400	1500	1300	1100	7600	138
139	1000	750	1000	750	800	1300	5600	138
140	700	800	600	600	800	1100	4600	120
141	1000	1100	900	1200	1300	1150	6650	140
142	1000	1100	1200	1100	1100	1300	6800	142
143	1200	2000	1800	1600	1900	1600	10100	152
144	800	500	700	900	1000	1800	5700	120
145	700	800	800	700	600	1900	5500	122
146	1000	700	1000	1800	1300	2000	7800	120
147	900	750	800	1300	1600	1400	6750	140
148	400	750	1400	1300	1500	1800	7150	150
149	300	750	400	900	1000	1200	4550	140
150	600	800	900	800	1000	1000	5100	152
151	400	1000	1000	500	600	900	4400	120
152	500	700	800	1500	1600	1100	6200	124
153	700	1000	1800	800	1900	1800	8000	130
154	1000	1100	1600	1700	1800	1200	8400	150
155	1300	1200	1400	1800	1800	1000	8500	131
156	1200	850	1200	1400	1400	1200	7250	142
157	700	850	900	1400	1500	1300	6650	145
158	500	800	1600	1000	1100	1900	6900	138
159	750	1600	1000	3500	2100	2250	11200	201
160	750	1550	1500	1400	2400	2500	10100	192
161	750	700	1600	1250	2200	2250	8750	189
162	500	800	700	2700	2000	2750	9450	189
163	1500	1000	800	1200	2200	2250	8950	201
164	1750	1000	800	1250	2250	2250	9300	219
165	1500	1100	1000	1250	2000	2200	9050	207
166	1500	700	1600	1400	2100	2250	9550	222
167	1300	800	800	1500	2600	2500	9500	210
168	1250	850	700	2500	2000	2300	9600	228
169	1000	850	750	2250	2100	2400	9350	207
170	650	900	800	2500	2000	2800	9650	180
171	650	650	850	1500	2600	2250	8500	240
172	600	600	800	2250	2100	2500	8850	189
173	500	550	800	3050	3000	2200	10100	168
174	1000	700	750	1500	2000	2250	8200	207
175	800	1000	1000	2000	2100	2600	9500	204
176	1000	800	1100	1200	2200	2250	8550	201
177	1100	1100	1150	1750	2150	2200	9450	177
178	1000	1150	1100	1750	2200	2400	9600	168
179	700	1600	1000	1500	1600	2250	8650	171
180	700	900	1150	1100	2200	2600	8650	168

181	750	850	1500	1500	2100	2500	9200	168
182	750	700	1600	2000	1900	3000	9950	174
183	800	750	1800	1900	1800	2500	9550	189
184	850	900	1600	2000	1950	2750	10050	192
185	600	1000	1650	2000	1100	2750	9100	189
186	600	950	1000	2200	2000	2000	8750	183
187	1000	750	1100	2000	1800	2000	8650	186
188	750	1100	1500	2000	1900	2100	9350	192
189	850	1000	1000	2500	1950	2000	9300	186
190	800	950	950	2250	2000	2800	9750	180
191	800	850	950	1200	2200	2800	8800	180
192	800	1000	950	1700	1500	2250	8200	177
193	750	1500	850	1500	2600	2500	9700	186
194	500	450	500	500	900	1300	4150	100
195	700	350	400	500	700	1250	3700	120

Appendix C. Sample Computation

Simple Random Sampling (SRS)(for average)

N = 899

n = 195

$$\overline{Y}_n = \frac{\sum_{i=1}^{n} Y_i}{n} = \frac{1405700}{195}$$

$$= 7208.718$$

$$s^2 = \frac{\sum \left(Y_i - \overline{Y}_n\right)^2}{n-1} = 10{,}911{,}596$$

$$V\left(\overline{Y}_n\right) = \frac{(N-n)}{N}\left(\frac{s^2}{n}\right)$$

$$= \frac{899-195}{899}\left(\frac{10911596}{195}\right)$$

$$= 43{,}819.4209$$

SE ($\overline{Y}_n$) = $\sqrt{V(\overline{Y}_n)}$ = 209.331

$$\text{CV}\ (\overline{Y}_n) = \frac{SE\left(\overline{Y}_n\right)}{\overline{Y}_n}$$

$$= \frac{209.331}{7208.718}(100)$$

$$= 2.904$$

Confidence Interval of the mean estimate:

$$P\left[\overline{Y}_n - t_{\frac{\alpha}{2},n-1} SE\left(\overline{Y}_n\right) \le \overline{Y}_n \le \overline{Y}_n + t_{\frac{\alpha}{2},n-1} SE\left(\overline{Y}_n\right)\right] = 1-\alpha$$

$$= 7{,}208.718 \pm 1.96(209.331)$$

$$= 6{,}798.429 \le \overline{Y}_n \le 7{,}619.007$$

Simple Random Sampling(for total)

$N\overline{Y}_n$ = 899(7208.718)
= 6,480,637.482

$$\hat{V}(N\,\overline{Y}_n) = \frac{N(N-n)}{n}s^2$$

= 35,414,899,810

$$SE(N\overline{Y}_n) = \sqrt{V(\overline{Y}_n)}$$

= 188,188.469

Confidence interval of the total estimate:

$$P\left[N\overline{Y}_n - t_{\frac{\alpha}{2},n-1}SE\left(N\overline{Y}_n\right) \le N\overline{Y}_n \le N\overline{Y}_n + t_{\frac{\alpha}{2},n-1}SE\left(N\overline{Y}_n\right)\right] = 1-\alpha$$

= 6,480,637.482 ± 1.960 (188188.469)

= 6,111,788.082 $\le N\overline{Y}_n \le$ 6849486.881

Classical Ratio Estimation

$$\overline{Y}_R = \frac{\overline{Y}_n}{\overline{X}_n}\overline{X}_N = \frac{7208.718}{131.81}(180)$$

= 9844.239

$$\text{bias}(\overline{R}_n) = \left(\frac{N-n}{Nn\overline{x}_n^2}\right)\left\{\overline{R}_n s_x^2 - s_{xy}\right\}$$

$$= \left(\frac{899-195}{899(195)(131.81)^2}\right)\left[54.69(3,186.596) - 411.161\right]$$

= 0.0401

$$\overline{R}_n = \frac{\overline{Y}_n}{\overline{X}_n} = \frac{7208.718}{131.81}$$

= 54.69

$s_x^2 = \dfrac{\sum (x_i - \bar{x}_n)^2}{n-1} = 3{,}186.596$

$s_{xy}^2 = \dfrac{\sum (x_i - \bar{x}_n)(y_i - \bar{y}_n)}{n-1} = 169{,}053.579$

Bias $(\bar{Y}_R) = \bar{X}_N$ bias $(\bar{R}_n)$
= 180 (.0401)
= 7.218

$V(\bar{Y}_R) = \dfrac{N-n}{Nn(n-1)} \sum \left[y_i - \bar{R}_n x_i \right]^2$

$= \dfrac{899-195}{899(195)(194)} (2.08E09)$

= 43,056.623

$SE(\bar{Y}_R) = \sqrt{V(\bar{Y}_R)} = 207.501$

$MSE(\bar{Y}_R) = V(\bar{Y}_R) + \text{bias}((\bar{Y}_R))^2$
= 43,056.623 + 7.218^2
= 43,108.723

$CV(\bar{Y}_R) = \sqrt{\dfrac{MSE(\bar{Y}_R)}{Y_R}} 100 = \sqrt{\dfrac{43061.605}{9844.239}} (100)$
= 2.109

<u>Classical Ratio Estimation (Total)</u>

$N\bar{Y}_R$ = 899 (9844.239)
= 8,484,346.662

$V(N\bar{Y}_R) = N^2 V(\bar{Y}_R) = (899)^2 (43056.623)$
= 34,798,405,770

$SE(N\bar{Y}_R) = \sqrt{V(N\bar{Y}_R)} = 186548.308$

$\text{bias}(N\bar{Y}_R) = N(\text{bias}(\bar{Y}_R))$

= 899(7.218)
= 6,488.982

Confidence Interval of the mean estimate:

$$P\left[\overline{Y}_n - t_{\frac{\alpha}{2},n-1} SE\left(\overline{Y}_n\right) \le \overline{Y}_n \le \overline{Y}_n + t_{\frac{\alpha}{2},n-1} SE\left(\overline{Y}_n\right)\right] = 1-\alpha$$

$= 9,844.239 \pm 1.960(207.501)$

$= 9,437.537 \le \overline{Y}_n \le 10250.941$

Confidence Interval of the total estimate:

$$P\left[N\overline{Y}_n - t_{\frac{\alpha}{2},n-1} SE\left(N\overline{Y}_n\right) \le N\overline{Y}_n \le N\overline{Y}_n + t_{\frac{\alpha}{2},n-1} SE\left(N\overline{Y}_n\right)\right] = 1-\alpha$$

$= 8,484,346.662 \pm 1.960(186,543.308)$

$= 8,118,721.778 \le N\overline{Y}_n \le 8,849,971.546$

Hartley-Ross Ratio-Type Estimation(average)

$$\overline{r}_n = \frac{1}{n}\sum_{i=1}^{n} r\underline{i} = \frac{1}{195}(10844.39)$$

$= 55.612$

$$\text{Bias}(\overline{Y}_r) = \frac{N-1}{N}(s_{rx}) = \frac{899-1}{899}(-122.126)$$

$= 121.99$

$$S_{rx} = \frac{1}{n-1}\left[\overline{y}_n - \overline{r}_n\overline{x}_n\right]$$

$$= \frac{1}{195-1}\left[7208.718 - 55.612(131.81)\right]$$

$=-122.126$

$$\overline{y}_r = \overline{r}\overline{X}_N + \frac{(N-1)n}{N(n-1)}\left(\overline{y}_n - \overline{r}_n\overline{x}_n\right)$$

$$= 55.612(180) + \frac{(899-1)}{899(195-1)}\left(7208.718 - 55.62(131.81)\right)$$

$= 9,888.170$

$$\hat{V}\ (\bar{y}_r) = \frac{1}{n}\frac{(N-1)}{N}\left(S_y^2 - 2\bar{r}_N S_x^2\right) + \frac{1}{n(n-1)}\frac{(N-1)^2}{N^2}\left(S_{rx}^2 + S_r^2 S_x^2\right)$$

$$= \frac{1}{195}\frac{(899-1)}{899}\left(11{,}249{,}454.99 - 2(55.612)(411.161)\right) + \frac{1}{195(195-1)}\frac{(899-1)^2}{899^2}$$

$$\left(1220.102 + 689.582(3{,}186.596\right)$$

$$=57{,}449.827$$

$$S_x^2 = \frac{\sum\left(x_i - \bar{x}_n\right)^2}{n-1} = 3{,}186.596$$

$$S_y^2 = \frac{\sum\left(y_i - \bar{y}_n\right)^2}{n-1} = 11{,}249{,}454.99$$

$$S_r^2 = \frac{\sum\left(r_i - \bar{r}_n\right)^2}{n-1} = 1220.102$$

$$S_{xy}^2 = \frac{\sum\left(x_i - \bar{x}_n\right)\left(y_i - \bar{y}_n\right)}{n-1} = 169{,}053.580$$

$$SE(\bar{y}_r) = \sqrt{V\left(\bar{y}_r\right)} = 209.950$$

$$MSE(\bar{y}_r) = V\left(\bar{y}_r\right) + \left(Bias\left(\bar{y}_r\right)\right)^2 = 44070.356 + (121.99)^2$$

$$= 72{,}331.387$$

$$CV(\bar{y}_r) = \sqrt{\frac{MSE\left(\bar{y}_r\right)}{\bar{y}_r}}100 = \frac{72331.387}{9888.170}(100)$$

$$= 2.720$$

Confidence Interval of the Mean

$$P\left[\bar{Y}_r - t_{\frac{\alpha}{2},n-1} SE\left(\bar{Y}_r\right) \le \bar{Y}_r \le \bar{Y}_r + t_{\frac{\alpha}{2},n-1} SE\left(\bar{Y}_r\right)\right] = 1-\alpha$$

$$= 9{,}888.170 \pm 1.96(\ 239.687)$$

$$= 9{,}418.383 \le \bar{Y}_r \le 10{,}357.957$$

Hartley-Ross Ratio-Type Estimation(total)

N$\overline{Y}_r$ = 899(9,88.170)

V(N$\overline{Y}_r$) = N^2 V($\overline{Y}_r$) = 899^2 (9888.170)
= 4.643100763 E 10

SE(N$\overline{Y}_r$) = $\sqrt{V\left(N\overline{y}_r\right)}$ = 215,478.555

Bias (N$\overline{Y}_r$) = N (Bias ($\overline{Y}_r$)) = 899(121.99)
=109669.01

Confidence Interval of the Total

$$P\left[N\overline{Y}_r - t_{\frac{\alpha}{2},n-1} SE\left(N\overline{Y}_r\right) \leq N\overline{Y}_r \leq N\overline{Y}_r + t_{\frac{\alpha}{2},n-1} SE\left(N\overline{Y}_r\right)\right] = 1-\alpha$$

= 8,889,464.83 $\pm$ 1.96(215,478.555)

= 8,467,126.872 $\leq$ N$\overline{Y}_r$ $\leq$ 9,311,802.788

Regression Estimation(average)

$\overline{Y}_{reg} = \overline{y}_n + \beta\left(\overline{X}_N - \overline{x}_n\right)$ = 7208.718 + 29.144 (180-131.81)
= 8613.167

$$\beta = \frac{\sum\left(x_i - \overline{x}_n\right)\left(y_i - \overline{y}_n\right)}{\sum_{i=1}^{n}\left(x_i - x_n\right)^2} = \frac{5570391}{191136.1}$$

= 29.144

$$V(\overline{Y}_{reg}) = \left(1 - \frac{n}{N}\right)\frac{se^2}{n} = \left(1 - \frac{195}{899}\right)\frac{10103092.78^2}{195}$$

= 40572.587

$$\hat{\beta}_1 = \frac{n\sum x_i - y_i - \sum\left(x_i\right)\left(y_i\right)}{n\sum{}_i^2 - \left(\sum x_i\right)^2} = \frac{195(1.92E09) - (25703(1405700))}{195(3596429) - (25703)^2}$$

=32.201

e$_i$ = y$_i$ - $\overline{Y}_n + \hat{\beta}_1\left(x_i - \overline{x}_n\right)$ = 2962.6841

$$\bar{e}_n = \frac{1}{n}\sum e_i = \frac{1}{195}(2862.684)$$
$$= 15.193$$

$$s_e^2 = \frac{\sum\left(e_i - \bar{e}_n\right)^2}{n-1} = \frac{1.96E09}{194}$$
$$= 10103092.78$$

$$SE\left(\bar{Y}_{reg}\right) = \sqrt{V\left(\bar{Y}_{reg}\right)} = 40572.587$$

$$\text{bias}\left(\bar{Y}_{reg}\right) = \frac{1-\frac{n}{N}}{n_{sx}^2}\sum_{i=1}^{N}\frac{e_i\left(x_i - \bar{x}_n\right)^2}{N-1}$$

$$= \frac{1\frac{195}{899}}{195(985.238)}\left(\frac{12591.27(2962.6941) - 131.86}{898}\right)$$
$$= 0.0223$$

$$MSE\left(\bar{Y}_{reg}\right) = V\left(\bar{Y}_{reg}\right) + \left(Bias\left(\bar{Y}_{reg}\right)\right)^2$$
$$= 40572.587 + (\ .0718)^2$$
$$= 40572.568$$

$$CV\left(\bar{Y}_{reg}\right) = \frac{\sqrt{MSE\left(\bar{Y}_{reg}\right)}}{\bar{Y}_{reg}}100$$
$$= 40{,}572.568$$

Confidence Interval of the mean:

$$P\left[\bar{Y}_{reg} - t_{\frac{\alpha}{2,n-1}} SE\left(\bar{Y}_{reg}\right) \le \bar{Y}_{reg} \le \bar{Y}_{reg} + t_{\frac{\alpha}{2,n-1}} SE\left(\bar{Y}_{reg}\right)\right] = 1-\alpha$$
$$= 8{,}613.167 \pm 1.960(201.426)$$
$$= 8{,}218.372 \le \bar{Y}_{reg} \le 9{,}007.962$$

<u>Regression Estimation(total)</u>

$$\bar{Y}_N = \mathrm{N}\left(\bar{Y}_{reg}\right) = 899(8613.167)$$
$$= 7{,}743{,}237.133$$

$$\mathrm{V}\left(N\bar{Y}_{reg}\right) = \mathrm{N}^2\,\mathrm{V}\left(\bar{Y}_{reg}\right) = 899^2\,(40572.587)$$
$$= 3.27908922\ \mathrm{E}\ 10$$

$SE\left(N\overline{Y}_{reg}\right) = \sqrt{V\left(N\overline{Y}_{reg}\right)}$ = 181,082.272

$bias\left(N\overline{Y}_{reg}\right) = \mathrm{N}\left(bias(\overline{Y}_{reg})\right)$ = 899(0.0223)
= 20.048

Confidence Interval of the total

$$P\left[N\overline{Y}_{reg} - t_{\frac{\alpha}{2,n-1}} SE\left(N\overline{Y}_{reg}\right) \le N\overline{Y}_{reg} \le N\overline{Y}_{reg} + t_{\frac{\alpha}{2,n-1}} SE\left(N\overline{Y}_{reg}\right)\right] = 1-\alpha$$

= 7,743,237.133 $\pm$ 1.960 (181,082.317)

= 7,388,315.88 $\le N\overline{Y}_{reg} \le$ 8,098,158.386

Magnitude of Efficiency

RE(Classical ratio over "Y-only", in %) = $\frac{V\left(\overline{Y}_n\right)}{V\left(\overline{Y}_R\right)}(100)$

$= \frac{43819.421}{43056.623}(100)$

= 101.77

RE (Hartley-Ross over "Y-only", in %) = $\frac{V\left(\overline{Y}_n\right)}{V\left(\overline{Y}_r\right)}(100)$

$= \frac{43819.421}{57449.827}(100)$

= 76.274

RE (Regression over "Y-only", in %) = $\frac{V\left(\overline{Y}_n\right)}{V\left(\overline{Y}_{reg}\right)}(100)$

= 108.00