Available online at www.elixirpublishers.com (Elixir International Journal)

Statistics

Elixir Statistics 43 (2012) 6994-7003

On double sampling approach for comparing estimates of students' enrolment in Oyo state public secondary schools

O.M. Olayiwola¹, S.A. Agboluaje², O.O. Salawu² and J.O. Fantola² ¹Statistics Department, Federal University of Agriculture, Abeokuta, Ogun State, Nigeria.

²The Polytechnic Ibadan, Oyo State, Nigeria.

ARTICLE INFO

Article history: Received: 3 November 2011; Received in revised form: 20 February 2012; Accepted: 22 February 2012;

Keywords

Double sampling, Students' Enrolment, Average over time, Matched portion.

ABSTRACT

Successive sampling is used repeatedly to survey a population over time. It allows the first sample to be taken (on the first occasion) and a second sample is then taken (on the second occasion). The scheme provides an opportunity of making use of the information obtained in the first sample in improving on the precision of future estimate. In this study, double sampling for regression estimation was used to determine the current estimate of the mean, minimum variance, maximum precision, estimate of change between the two successive occasions under consideration and estimate of average over the period of the two occasions.

The data used were based on students' enrolment in Oyo state public secondary schools and The data were collected from Planning, Research and Statistics Unit, Oyo State Ministry of Education. The current estimates for the student enrolments in Junior and Senior Secondary Schools were found to be 6,372 and 4,925 students respectively. The variances for the current estimate is more precise when ρ approaches unity. Sensitivities increased for both the Junior and Senior secondary schools at $0.7 \le \rho \le 1.0$. The changes in current estimates between the first and second occasion were found to be -110 and -210 for Junior and Senior Secondary Schools respectively. The negative sign shows that there was decreased in number of the students enrolment in both Junior and Senior Secondary Schools in the

current year compared with the previous year. Sensitivity was more in $\mu - \rho$ for $\mu \le 0.32$

and in μ for $\rho > 0.9225$ and $\rho > 0.7385$ for Junior and Senior Secondary Schools respectively. The estimate of average over time for the student's enrolment in Junior and Senior Secondary were found to be 12, 823 students and 10,063 students respectively. The gain in information changing from one occasion to the next was 238.98% for junior category and 45.37% for Senior Category.

© 2012 Elixir All rights reserved.

Introduction

The theory of successive sampling appears to have started with the work of Jessen (1942). He utilized the entire information collected in the previous occasions and obtained two estimates, one was the sampling mean based on new sample only and the other was a regression estimate based on the sample units observed on both occasions by combining the two estimates.

Yates (1949) extended Jessen's scheme to the situation where the population mean of a variable is estimated on each one of the h (≥ 2) occasions from a rotational sample design. The results were generalized by Patterson (1950) and Tikkiwal (1951). Patterson assumed that the correlation between observations on the same units on occasion *h* and occasion *h* + k is p^k (h, k= 1, 2...,). He arrived at this using partial replacement of units from occasion to occasion. A set of conditions were established in the work whereby a proposed estimate may be tested for efficiency. It was shown that in the case in which correlation coefficient ρ is available and the number of units replicated is the same for all occasions, efficiency estimates are not easily calculated for the last but one occasion. Patterson (1950) aimed at providing the optimum estimate by combining (i) a double sample regression estimate from the matched of the values of y on the first occasion and (ii) a sample based on a random sample from the unmatched portion on the second occasion. The theory was generalised to provide the optimum estimate using p auxiliary variables (p>1).

Eckler (1955) introduced the idea of rotation sampling and obtained the minimum variance unbiased estimator (MVUE) of the population mean assuming an infinite population.

Tikkiwal (1955) developed the estimation of the mean of several characters in a multipurpose sampling on occasions. He assumed the pattern of the correlation between the ith and the jth occasion to be of the form. $p_{ij}^{ih} = p_{ij}^{i} \cdot p_{j}^{ih}$

Rao (1957) considered the estimate of the population ratio based on sampling two occasions. He used a ratio type estimator of the form

$$\hat{R} = r_2 R_1 / r_1$$

Where r_1 and r_2 are sample ratio of two characters for the first and second occasions respectively. R is known as the population ratio of the first occasion.



Kulldorf (1963) modified Jessen's scheme of the sampling by selecting the unmatched sample from the units not selected on the first occasion. He considered in detail the optimum choice of the matching fraction under the most general form of cost function apart from fixed costs.

Rao and Grahim (1964) developed a unified population theory for composite estimators and employed fixed rotation design in a finite population.

Raj (1965) pioneered the use of varying probability with replacement for sampling over two successive occasions by using probability proportional to size with replacement (PPSWR) and a simple random sampling without replacement (SRSWOR). His estimation for the population total on the second occasion was based on a linear combination of two independent estimates of the population total from the matched and unmatched samples. This Raj's estimator was modified by Pathak and Rao(1967). Ravindra Singh (1972, 1980) modified Raj's estimator further.

Tikkiwal (1967) again dealt with the study of k characters in a multipurpose survey for finite and infinite populations on each h>2 occasions under a specific correlated pattern.

Singh, D (1968) was the first person who extended the theory of one-stage sampling to two-stage sampling scheme over two occasions. His scheme sampling is as follows:

On the first occasion, a sample of n first stage units (F.S.U) is selected by simple random sampling and without replacement (SRSWOR) from a population of N.F.S.U. Each of the selected F.S.U was sub-sampled by SRSWOR method.

On the second occasion, a simple random sample np (0 of these F.S.U are retained along with their second stage unit (S.S.U) selected on the first occasion. A fresh sample of nq first stage unit <math>(p+q) = 1 is selected from the entire population again by SRSWOR to obtain

the required number of second stage unit.

Singh D (1968) obtained a minimum variance linear unbiased estimator for the population means assuming an infinite population. Abrahim, khosta and kathuria (1969) considered the sampling scheme where partial matching of units was carried out at both stages. Their result was applied to a survey on the incidence of pest and diseases.

Ghangurde and Rao (1969) extended Raj's strategy by introducing varying probability without replacement scheme of sampling for the succession of initial sample and unmatched sample S_2 .

Sen (1971) developed the theory of sampling over two occasions further using two auxiliary variables with unknown population mean. Sen(1972) generalised Jessen's work by using a double sampling multivariate ratio estimate using P-auxiliary variables (p>1) from the matched portion of the sample. Expressions for optimum matching fraction and the combined estimate and its error have also been derived. He also considered the theory of the case where the mean of matched sample on second occasion is adjusted by the multivariate ratio estimate for equal sample sizes.

Kathuria and Singh (1971) examined empirically the relative efficiencies of the three sampling schemes for sampling on successive occasions using two-stage design.

Shivtar and Srivastava (1973) improved on Singh's (1968) estimators by using several auxiliary variables in estimating the population mean on the most recent occasion, difference in mean between any two successive occasions and the overall population mean for all the occasions.

Srivastava and Shivtar (1974) used a more general two stage sampling scheme of partial retention of both primary units

and secondary units in line with Abraham et al (1968) auxiliary variable.

Tripathi and Sinha (1976) considered estimation of the population ratio over two occasions by taking a linear combination of the estimates of the population ratio based on matched and unmatched samples.

Arnab (1976) also generalized Raj (1965), Ghanguarde and Rao (1969) strategies to h>2 occasions. Tewari (1981) estimated the domain mean using successive sampling scheme.

Lokesh Arora and Singh (1981) extended the theory of successive sampling for the estimation of the population mean to the estimate of the frequency distribution of the population.

Arnab (1982) has shown that strategies due to Raj (1979) can be improved upon by selecting the current unmatched sample from the complements of the matched sub-sample for the current occasions.

Das (1982) extended Tripathi and Sinha's work using a single auxiliary variable for the estimation of the population ratio over two occasions.

Chaturvedi and Tripathi (1983) extended the theory further using more than one auxiliary variables.

Okafor (1987) compared some estimators of the population total in two-stage successive sampling using an auxiliary variable. Also Okafor (1988) obtained ratio-to-size estimators of the mean per subunit using two-stage sampling over two occasions. The estimation of the population ratio in two-stage sampling over two occasions was considered by Okafor and Arnab (1987).

Iuchan and Jones (1987) looked into rotation sampling patterns while Steel and Maclauren (2000) examined the effect of these rotations of trend estimates.

Artes, Rueda and Arcos (1998), in examining the relationship between the auxiliary variables X and the study variable Y, proved that when X and Y are negative, the optimum estimate which combines a double sample product estimates for the matched part of the sample and simple sample mean for the unmatched portion has less variance than the usual estimator

provided that,
$$\rho > -\frac{C_x}{2C_y}$$
.

Artes and Garcia (2001) worked on estimating the current mean in successive sampling using a product estimate.

Garcia Luengo (2004) considered the problem of estimation of a finite population mean and for the current occasion based on the sample selected over two occasions for the case when, several auxiliary variables are correlated with the main variable. A double sampling multivariate product estimate from the matched portion of the sample is presented. Expressions for optimum estimator and its error have been derived. The gain in efficiency of the combined estimate over the direct estimate using no information gathered on the first occasion was computed.

Artes et al (2005) also worked on successive sampling using a product estimate but, they considered the case when the auxiliary variables are negatively correlated and double sampling product estimate from the matched portion of the sample was presented. Expression for optimum estimator and its variance have been desired.

Reuda et al (2006) talked on estimating quantitative under sampling on two occasions with arbitrary sample designs.

Rueda et al (2007) further extended the work on successive sampling in estimating quartiles with p-auxiliary variables. They mainly discussed the estimation of quartiles for the current occasion based on sampling on two occasions and using pauxiliary variables obtained from the previous occasions. A multivariate ratio estimate from the matched portion was used to provide the optimum estimate of a quartile by weighing the estimate inversely to derive optimum weight.

Housila et al (2007) looked into the problem of estimating a finite population quantile in successive sampling on two occasions. They aimed at providing the optimum estimates by combining (i) three double sampling estimator viz. ratio-type, product-type and regression-type estimator, from the matched portion of the sample and (ii) a sample quantile based on a random sample from the unmatched portion of the sample on the second occasion. A simulation study was carried out in order to compare the three estimators and it is found that the performance of the regression-type estimator is the best among all the estimators discussed.

Manish. and Shukla (2008) looked into the efficient estimator in successive sampling using post stratification. They stated that it is often seen that a population having large number of elements remains unchanged in several occasions but the value units change. In their work, they introduced an estimator under successive survey, the estimator is unbiased and efficient over post-stratification estimations, the minimum variance of the optimum estimator was derived and comparative study incorporated.

Housila et al (2010) looked into estimation of population variance in successive sampling and proposed a class of estimators of finite population variance in successive sampling on two occasions and analysed its properties. These classes of estimators were motivated by Isaki (1983) to consider the problem of estimation of finite population variance in survey sampling. A general class of estimators for estimating the finite population variance σ^2_y based on the matched portion of the sample consisting of m units is defined as $S_t^2 = t(S_{ym}^2, v_1, v_2)$, the bias and variance of the estimator, \mathbf{S}_{t}^{2} , exist since the number of possible samples is finite and it is assumed that the function is bounded. The bias of the estimator S_{\star}^{2} is of the order of m^{-1} and hence its contribution to the mean square error will be of the order m^{-2} . The new estimator S^2 proposed which is more efficient than the usual unbiased estimator $S^2_{y_2}$ was gotten from the combination of a bias and unbiased estimator based on unmatched units.

The intention of this paper is therefore to compare the estimates of students' enrolment in public junior and secondary schools in Oyo state.

The data employed for this study is a secondary data collected from Planning, Research and Statistics Unit, Oyo State Ministry of Education.

Methodology

A population sampled over two occasions is considered for making current estimates of the population characteristics. On the first occasion, a sample of n units is selected from N units in the population by SRSWOR.

On the second occasion a sub-sample of λn units, $(0 < \lambda < 1)$, is selected from the first occasion sample by SRSWOR. This is supplemented by a fresh sample of μn units $(\mu + \lambda = 1)$, selected by SRSWOR from N or from (N-n) units. Information on the characters X and Y are obtained on all the sample units on both occasions.

Current Estimates Of The Mean And Variance

$$\hat{\mu}_{1} = \frac{1}{n} \sum_{i=1}^{n} x_{i} = \bar{x}, \qquad \hat{\mu}_{2u} = \frac{1}{n\mu} \sum_{i=1}^{m\mu} y_{i} = \bar{y}^{11}$$
$$\hat{\mu}_{2m} = \bar{y}^{1} + b(\bar{x} - \bar{x}^{1})$$

Where b is regression coefficient between x and y

$$V\left(\hat{\mu}_{2u}\right) = \frac{\sigma^2}{n\mu}$$

 $n\mu$, Where σ^2 is the variance of the variate at second occasion.

correlation coefficient between x and y (i.e 1^{st} and 2^{nd} occasions).

Estimation of optimum variance

$$V_0\left(\stackrel{\wedge}{\mu}_2\right) = \frac{\sigma^2}{2n} \left[1 + \sqrt{2\left(1 - \stackrel{\wedge}{\rho}\right)}\right]$$

The equation attains its optimum when $\rho = 1$ The Minimum Variance Unbiased Estimator (MVUE)

$$V_0\left(\hat{M}_2\right) = \frac{\sigma^2}{2n} \left[1 + \left(1 - \hat{\rho}^2\right)^{\frac{1}{2}}\right]$$

Where ρ is correlation between 1st and 2nd occasion and $\begin{pmatrix} \land \\ \end{pmatrix}$

 $V_0\left(\stackrel{\wedge}{M}_2\right)$ attains its optimum when $\rho = 1$.

Estimates of Change

To estimate change from one occasion to another; we took the difference between the previous and current estimates.

$$V\left(\hat{\Delta}\right) = \frac{2\sigma^2\left(1-\hat{\rho}\right)}{n\left(1-\mu\,\hat{\rho}\right)}$$

We have the optimum variance when the $\mu = 0$, $\rho = 1$ i.e. perfect matching of the same points on both occasions under estimation of change.

Average over Time

To obtain this, we require the sum of the estimators over both occasions.

Let the sum =
$$\varepsilon = M_1 + M_2 \Longrightarrow \overset{n}{\varepsilon} = \overset{n}{M_1} + \overset{n}{M_2}$$
 and the average over time = $\frac{\overset{n}{\varepsilon}}{\overset{n}{\varepsilon}}$.

Variance of Average over Time

2

$$V\left(\hat{\varepsilon}\right) = \frac{2\left(1+\hat{\rho}\right)\sigma^{2}}{\left(1+\mu\,\hat{\rho}\right)n}$$

We attained the minimum variance when $\rho = 0$, $\mu = 1$, which means taking independent samples enhances better estimates of average over time.

The Gain from Common Samples

The precision of estimation can be determined by evaluating the efficiency of the relative gain in information from the sample from one occasion to the next.

The mean on the first occasion is, $\lambda x^{-1} + \mu x^{-11}$ while that for the second occasion is; $\lambda y^{-1} + \mu y^{-11}$. We may eliminate the common samples by taking the differences between the two samples.

R.E =
$$\frac{\lambda \mu \rho^2}{1-\rho} x \frac{100}{1}$$
, since $\lambda + \mu = 1$

The gain from common sample increases as $\rho \sim 1$. However, it seems equal precision may be achieved by keeping the same sample or changing it from occasion to occasion. Data Analysis And Discussion

Data Analysis And Discussion

A random sample of size 25 is selected on each occasion, these comprised 17 matched and 8 unmatched samples. i.e. n = 25

$$m = n\lambda = 17$$

$$25\lambda = 17$$

$$\lambda = 0.68$$
But $\mu + \lambda = 1$

$$\mu = 1 - 0.68$$

$$\mu = 0.32$$

The estimates in thousands are given as follows.

Estimates on The Number of Students Environment in Public Junior Secondary Schools

Current estimates of mean and variance for Junior Category

 $\hat{\mu}_{2m} = \overline{y}^{1} + b\left(\overline{x} - \overline{x}^{1}\right)$ From the data, we have

$$\Sigma x = 161.514 ; \qquad x = 6.461$$

$$\Sigma y = 135.855 ; \qquad \overline{y} = 5.434$$

$$\Sigma x^2 = 1,738.974 ; \qquad \Sigma xy = 939.255$$

$$\Sigma y^2 = 1,100.218 ;$$

Recall that

$$b = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{n\Sigma x^2 - (\Sigma x)^2}$$
$$= \frac{25(939.255) - (161.514)(135.855)}{25(1,738.974) - (161.514)^2}$$

b = 0.0885Also from the data, we have

$$\Sigma x^{1} = 88.221 \qquad ; \qquad \Sigma y^{2} = 582.3318$$

$$\Sigma y^{1} = 85.908 \qquad ; \qquad \overline{y} = 5.0534$$

$$\Sigma x^{1} y^{1} = 583.6516 \qquad ; \qquad \overline{x} = 5.1894$$

$$\Sigma x^{12} = 609.1115$$

$$\therefore \hat{\sigma}^{2} = \frac{1}{n} \left[\Sigma y^{2} - n\overline{y}^{2} \right]$$

$$= \frac{1}{25} \left[(1,100.218) - 25(5.434)^{2} \right]$$

$$\hat{\sigma}^{2} = 14.4803$$
(a) $\hat{\mu}_{2m} = \overline{y}^{-1} + b(\overline{x} - \overline{x}^{-1}) = 5.0534 + 0.0885(6.461 - 5.1894) = 5.0534 + 0.1125 = 5.1659$

(b)
$$\hat{\mu}_{2u} = \overline{y}^{-11} = \frac{\Sigma y^{11}}{n\mu} = \frac{65.387}{8} = 8.1728$$

(c)
$$V\left(\stackrel{\wedge}{\mu}_{2u}\right) = \frac{\stackrel{\wedge}{\sigma}}{n\mu} = \frac{14.4803}{8} = 1.8100$$

(d)
$$V\left(\stackrel{\wedge}{\mu}_{2m}\right) = \frac{\sigma^2}{n\lambda} \left[1 + (1 - \lambda)(1 - 2\stackrel{\wedge}{\rho})\right]$$

Where
$$\hat{\rho} = \frac{n2xy - 2x2y}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$$

 $\hat{\rho} = \frac{17(583.9494) - (88.221)(85.905)}{\sqrt{[17(609.1115) - (88.221)^2][17(582.3318) - (85.908)^2]}}$
 $\hat{\rho} = \frac{2348.2501}{2545.5681}, \quad \hat{\rho} = 0.9225$
 $V_{\min}(\hat{\mu}_{2m}) = \frac{14.4803}{17}[1 + (0.32)(1 - 2(0.9225))] =$
14.4803(0.7296) = 0.6215
 $V_{\min}(\hat{M}_2) = \frac{\sigma^2}{2n} \left[1 + \left(1 - \hat{\rho}^2\right)^{\frac{1}{2}}\right] =$
 $\frac{14.4803}{50} \left[1 + (0.1490)^{\frac{1}{2}}\right] = \frac{14.4803}{50}(1.3860)$
 $\hat{V}_{\min}(\hat{M}_2) = 0.4014$

Test for current estimates of variance with different values of $\stackrel{\wedge}{\rho}$

The current estimate is function of $\hat{\rho}$ only. Estimates of changes for junior category

From the data, we have

$$\Sigma x^{11} = 60.509 \qquad ;$$

$$\overline{x}^{11} = 7.5636$$

$$\Sigma y^{11} = 65.382 \qquad ;$$

$$\overline{y}^{11} = 8.1728$$

$$\widehat{M}_{2} = \frac{1}{1 - \mu^{2} \rho^{2}} \left[\lambda \mu \rho \left(\overline{x}^{11} - \overline{x}^{1} \right) + \lambda \overline{y}^{1} + \mu \left(1 - \mu \rho^{2} \right)^{2} \right]$$

Where,
$$\mu = 0.32, \lambda = 0.68, \rho = 0.9225$$

$$\overline{x}^{1} = 5.1894, \overline{y}^{1} = 5.0534$$
$$\frac{1}{1 - \mu^{2} \rho^{2}} = \frac{1}{1 - [0.32(0.9225)]^{2}} = 1.0955$$

=

$$\lambda \mu \hat{\rho} \begin{pmatrix} z^{-11} & -z^{-1} \\ z^{-1} & -z^{-1} \end{pmatrix} = 0.68(0.32)(0.9225)[7.5636 - 5.1894] = 0.4766$$

$$\lambda \overline{y}^{-1} = 0.68(5.0534) = 3.4363$$

$$\mu \left(1 - \mu \hat{\rho}^{-2}\right) \overline{y}^{-11} = 0.32[1 - (0.32(0.9225)^2)] 8.1728 = 1.9032$$

Then

I nen

$$\hat{M}_2 = 1.0955(0.4766 + 3.4363 + 1.9032)$$

1.0955(5.8161) = 6.3715

The current estimate of number of students' enrolment in junior category per local government is 6.3715 x 1000 \approx 6,372 students.

$$M_{1} = \frac{1}{1 - \mu^{2} \rho^{2}} \left[\lambda \mu \dot{\rho} \left(y^{-11} - y^{-1} \right) + \lambda x^{-1} + \mu \left(1 - \mu \dot{\rho}^{2} \right) x^{-11} \right] \text{Thus}, \qquad \text{eff}$$

$$\frac{1}{1 - \mu^{2} \rho^{2}} = 1.0955$$

$$\lambda \mu \dot{\rho} \left(y^{-11} - y^{-1} \right) = 0.68(0.32)(0.9225)[8.1728 - 5.0534] = 0.6\overline{26}$$

$$\lambda x^{-1} = 0.68(5.1894) = 3.5288 \qquad \mu \left(1 - \mu \dot{\rho}^{2} \right) x^{-11} = 0.32[1 - 0.32(0.9225)^{2}]7.5636 = 1.7613 \qquad 9.6$$

Then

 $M_1 = 1.0955(0.6262 + 3.5288 + 1.7613) = 1.0955$ (5.9163) = 6.4813

The previous estimate of number of students' enrolment in junior category per local government is 6.4813 x 1000 \approx 6,482 students

Estimate of change;

 $\Delta = M_2 - M_1 = 6.3715 - 6.4813 = -0.1098$ The change in current estimate between the first and second occasion is $-0.1098 \ge 1000 = -110$ students

$$\hat{V}(\hat{\Delta}_{1}) = \frac{2(1-\hat{\rho})\hat{\sigma}^{2}}{(1-\mu\hat{\rho})n} = \frac{2(1-0.9225)14.4803}{1-0.32(0.9225)25}$$
$$= \frac{2.2445}{17.65} = 0.1274$$
$$\hat{V}(\hat{\Delta}_{1}) = 0.1274$$

Relative Gain For Junior Category

$$\frac{\hat{V}\left(\hat{\Delta}_{2}\right)}{\hat{V}\left(\hat{\Delta}_{1}\right)}\!-\!1$$

Where
$$\hat{V}(\hat{\Delta}_2) = 2(1 - \lambda \hat{\rho})\frac{\hat{\sigma}^2}{n} = 2[1 - 0.68(0.9225)]\frac{14.4804}{25}$$

 $\hat{V}(\hat{\Delta}_2) = 0.4318$

Relative gain for junior category is = $\left[\frac{0.4318}{0.1272} - 1\right] x100$ =

2.3893 x 100 = 238.98%

The above table shows that substantial gains in precision are achieved by using the better estimator when $\rho = 1$ and the value of $\lambda = 0.5$ are unique and show maximum precision. Estimates of average overtime for juni

Estimates of average overtime for junior category

$$\hat{\Delta} = \frac{1}{1+\mu\rho} \left[\mu \left(1+\rho \right) \left(y^{-1} + x^{-11} \right) + \lambda \left(y^{-1} + x^{-1} \right) \right]$$

$$\frac{1}{524} \frac{1}{1+\rho\rho} = \frac{1}{1+0.32(0.9225)} = 0.7721$$

$$\mu \left(1+\rho \right) \left(y^{-11} + x^{-11} \right) = 0.32(1+0.9225)(8.1728+7.5636)$$
9.6810

$$\lambda \left(y^{-1} + x^{-1} \right) = 0.68(5.1894+5.0534) = 6.9651$$

=

$$\hat{\Delta} = 0.7721 \ (9.6810 + 6.9651) = 12.8525$$

$$\hat{V}\left(\hat{\Delta}\right) = \frac{2\left(1 + \hat{\rho}\right)\hat{\sigma}^2}{\left(1 + \mu\hat{\rho}\right)n} = \frac{2(1 + 0.9225)(14.4803)}{[1 + 0.32(0.9225)](25)} = \frac{55.6768}{32.38}$$

$$\hat{V}\left(\hat{\Delta}\right) = 1.7195$$

Estimates on number of students' enrolment in public senior secondary schools

Current Estimates of mean and variance for Senior Category

$$\Sigma x = 151.022 \quad ; \qquad \overline{x} = 6.0409$$

$$\Sigma y = 102.272 \quad ; \qquad \overline{y} = 4.0909$$

$$\Sigma xy = 579.9437 ; \Sigma y^2 = 663.9307$$

$$\Sigma x^2 = 2,334.7842$$

$$b = \frac{25(579.9437) - (151.022)(102.272)}{25(2,334.7842) - (151.022)^2} , \qquad b = -0.0266$$

$$\overset{\wedge}{\sigma}^2 = \frac{1}{n} \left[\sum_{i=1}^n y^2 - n\overline{y}^2 \right] \qquad = 9.8219$$

$$\frac{1}{25} \left[663.930 - 25(4.0909)^2 \right] = 9.8219$$

$$\Sigma x^1 = 70.591 \quad ; \qquad \overline{x} = 4.1524$$

$$\Sigma y^1 = 62.997 \quad ; \qquad \overline{y} = 3.7057$$

$$\Sigma x^1 y^1 = 30633 \quad ; \qquad \Sigma y^{1^2} = 296.1327$$

$$\Sigma x^{1^{2}} = 351.6806$$
(a) $\hat{\mu}_{2m} = \overline{y}^{1} + b(\overline{x} - \overline{x}^{1}) = 3.7057 - 0.0266(6.0409 - 4.1524) = 3.7057 - 0.0502 = 3.6555$
(b) $\hat{\mu}_{2m} = \overline{y}^{-11} = \overline{\sum y^{11}} = \frac{57.838}{57.838} = 7.230$

(b)
$$\mu_{2\mu} = y = \frac{2}{n\mu} = \frac{2}{8} = 7.230$$

(c)
$$\hat{V}\left(\hat{\mu}_{2\mu}\right) = \frac{\hat{\sigma}}{n\mu} = \frac{9.8219}{8} = 1.2277$$

(d)
$$\hat{V}\left(\hat{\mu}_{2m}\right) = \frac{\hat{\sigma}^2}{n\lambda} \left[1 + (1 - \lambda)\left(1 - 2\hat{\rho}\right)\right]$$

where

$$\hat{\rho} = \frac{17(306.33) - (70.591)(62.997)}{\sqrt{[17(351.6806) - (70.591)^2][17(296.1327) - (62.997)^2]}} \frac{760.59}{\sqrt{[1029.96]}}, \quad \hat{\rho} = 0.7385$$

$$= \hat{V}(\hat{\mu}_{2m}) = \frac{9.8219}{17} [1 + (0.32)(1 - (0.7385))] = 0.4896$$
(e) $\hat{V}_{\min}(\hat{M}_2) = \frac{\sigma^2}{2n} [1 + (1 - \hat{\rho}^2)^{\frac{1}{2}}] = 0.4896$
(e) $\hat{V}_{\min}(\hat{M}_2) = \frac{\sigma^2}{2n} [1 + (1 - \hat{\rho}^2)^{\frac{1}{2}}] = \frac{16.44444}{50}, \quad \hat{V}_{\min}(\hat{M}_2) = 0.3289$

Estimates of change for senior category From data, we have

$$\Sigma x^{11} = 44.13$$
 ; $x^{-11} = 5.5163$
 $\Sigma y^{11} = 57.838$; $y^{11} = 7.2398$

When $\mu = 0.32$, $\lambda = 0.68$, $\rho = 0.385$

$$\vec{y} = 3.7057; \quad \vec{x} = 4.1524$$

$$\hat{M}_{2} = \frac{1}{1 - \mu^{2} \rho^{2}} \left[\lambda \mu \rho \left(\vec{x} - \vec{x} \right) + \lambda \vec{y} + \mu \left(1 - \mu \rho^{2} \right)^{-11} \right]$$

$$\frac{1}{1 - \mu^{2} \rho^{2}} = \frac{1}{1 - (0.32)^{2} (0.77385)^{2}} = 1.0592$$

$$\lambda \mu \rho \left(\vec{x} - \vec{x} \right) = 0.68 (0.32) (0.7385) [5.5163 - 4.1524]$$

$$= 0.1607 (1.3639) = 0.2192$$

$$\lambda \vec{y}^{1} = 0.68 (3.7057) = 2.5199$$

$$\mu \left(1 - \mu \dot{\rho}^2\right) \overline{y}^{-11} = 0.32 \left[(1 - 0.32)(0.7382)^2 \right] 7.2298 = 1.9101$$

Then,

$$\hat{M}_2 = 1.0592[0.2192 + 2.5199 + 1.9101] = 1.0592$$

(4.6492) = 4.9244

The current estimate of number of students' enrolment in senior category per local government is $4.9244 \times 1000 = 4,925$ students.

$$\hat{M}_{1} = \frac{1}{1 - \mu^{2} \rho^{2}} \left[\lambda \mu \rho \left(y^{-11} - y^{-1} \right) + \lambda x^{-1} + \mu \left(1 - \mu \rho^{2} \right) x^{-11} \right]$$
$$\frac{1}{1 - \mu^{2} \rho^{2}} = 1.0592$$
$$\lambda \mu \rho = \left(y^{-11} - y^{-1} \right) =$$
$$\lambda \mu \rho = (0.32)(0.7385)[7.2298 - 3.7057] = 0.1607(3.5241) = 0.5663$$

$$(0.32)(0.7385)[7.2298 - 3.7057] = 0.1607(3.5241) = 0.566$$
$$\lambda x^{-1} = 0.68(4.1524) = 2.836$$

$$\mu \left(1 - \mu \rho^{2}\right)^{-11} = 0.32 \left[1 - 0.32 (0.7385)^{2}\right] 5.5163 = 1.4572$$

Then,

$$\hat{M}_1 = 1.0592[0.5663 + 2.8236 + 1.4572] =$$

1.0592(4.8471) = 5.1341

The previous estimate of number of students' enrolment in senior category per local government is $5.1341 \times 1000 = 5,135$ students

Estimate of change;
$$\hat{\Delta} = \hat{M}_2 - \hat{M}_1$$

 $\Delta = 4.9244 - 5.1341 = -0.2097$

The change in current estimate between the first and second occasions is $-0.2097 \times 1000 = -210$ students.

$$\hat{V}(\hat{\Delta}_{1}) = \frac{2(1-\hat{\rho})\hat{\sigma}^{2}}{(1-\mu\hat{\rho})n} = \frac{2(1-0.7385)9.8219}{[1-0.32(0.7385)25]}$$
$$= \frac{5.1369}{19.092} = 0.2691$$
$$\therefore \hat{V}(\hat{\Delta}_{1}) = 0.2691$$

Relative gain for senior category

$$\frac{\hat{V}\left(\hat{\Delta}_{2}\right)}{\hat{V}\left(\hat{\Delta}_{1}\right)} - 1$$

Where $\hat{V}\left(\hat{\Delta}_{2}\right) = 2\left(1 - \lambda \hat{\rho}\right)\frac{\hat{\sigma}^{2}}{n}$

_

$$= 2[1 - 0.68(0.7385)]\frac{9.8219}{25}$$
$$\hat{V}(\hat{\Delta}_2) = 0.3912$$

. Relative gain for senior category is

$$= \left\lfloor \frac{0.3912}{0.2691} - 1 \right\rfloor x 100 = 0.4537 \text{ x } 100 = 45.37\%$$

Estimate of Average over time for senior category

$$\hat{\Delta} = \frac{1}{1+\mu\rho} \left[\mu \left(1+\rho\right) \left(\overline{y}^{11}+\overline{x}^{11}\right) + \lambda \left(\overline{y}^{1}+\overline{x}^{1}\right) \right] \\ \frac{1}{1+\mu\rho} = \frac{1}{1+0.32(0.7385)} = 0.8089 \\ \mu \left(1+\rho\right) \left(\overline{y}^{11}+\overline{x}^{11}\right) = 0.32(1+0.7385)(5.5163+7.2398) \\ 7.0962 \\ \lambda \left(\overline{y}^{1}+\overline{x}^{1}\right) = 0.68(4.1524+3.7057) = 5.3435 \\ \hat{\Delta} = 0.8089(7.0962+5.3435) = 10.0625 \\ \hat{V} \left(\hat{\Delta}\right) = \frac{2\left(1+\rho\right)\hat{\sigma}^{2}}{\left(1+\mu\rho\right)n} \\ = \frac{2(1+0.7385)9.8219}{\left[1+0.32(0.7385)\right]25} = \frac{34.1508}{30.909} = 1.1049$$

Conclusion

The current estimates for the student enrolments in Junior and Senior Secondary Schools were found to be 6,372 students and 4,925 students respectively. The variances for the current estimate is more precise when ρ approaches unity. Sensitivities increased for both the Junior and Senior secondary schools at 0.7 $\leq \rho \leq 1.0$.

The changes in current estimates between the first and second occasion were found to be -110 students and -210 students for Junior and Senior Secondary Schools respectively. The negative sign shows that there was decreased in number of the students enrolment in both Junior and Senior Secondary Schools in the current year compared with the previous year.

Sensitivity was more in $\hat{\rho}$ for $\mu \leq 0.32$ and in μ for $\hat{\rho} >$

0.9225 and $\rho > 0.7385$ for Junior and Senior Secondary Schools respectively.

The estimate of average over time for the student's enrolment in Junior and Senior Secondary were found to be 12, 823 students and 10,063 students respectively.

The gain in information changing from one occasion to the next was 238.98% for junior category and 45.37% for Senior Category.

References

Artes, E., Reuda M., and Arcos A.(1998): Successive sampling using a product estimate, applied sciences and the environment. *Computational mechanics publication* 85-90.

Artes, E. and Garcia A.V. (2001): Estimating the current mean in successive sampling using a product estimate. *Conference on agricultural and environmental statistical application in Rome*, XLIII-1 - XLIII-2.

Artes, E., Rueda M. and Arcos A.(2005): Successive sampling using a product estimate when the auxiliary variables are negatively correlated. *Applied science and environment*, page 6.

Chaturvedi, D.K., Tripathi T. P. (1983): Estimation of population ratio on two occasions using multivariate auxiliary information. *Journal Indian Stat. Assoc.* 21, 113-120.

Cochran, W.G (1977): *Sampling techniques (Third Edition)* John Wily and Sons (New York).

Das, A.K. 1982: Estimation of population ratio on two occasions. *Journal Indian Soc. Agric.Stat.* 34,1-9.

Das, A.K., Tripathi T.P.(1987): Use of auxiliary information in estimating the finite population Variance. *Sankhya* 40,139-148.

Eckler, A.R. (1955): Rotation sampling. Ann. Math. Stat.26, 664-685.

Ghangarde P.D., Rao J.N.K.1969: Some result on sampling over two occasions. *Sankhya* A31, 463-472.

Housila, P.S., Ritech T., Sarjinder S. and Jong-Min, K. (2007): Quantile estimation in successive sampling. *Journal of the Korean statistical society* 36 (4) 543-556.

Housila, P.S., Ritech T., Sarjinder S. and Jong-Min, K. 2010: Estimation of Population variance in successive sampling.*cda.mrs.Umn.edu/jongmink/research*.

Jessen, R.J. 1942: *Statistical investigation of a sample survey for obtaining farm facts*. Lowa agriculture experiment statistical research bulletin, 304.

Kullduf, G. 1963: Some problem of optimum allocation for sampling on two occasions. *Review International statistical institute*, 31, 24-57.

Kathuria, O.P, Singh D. 1971: Comprising of estimates in two stage sampling on successive occasions. *Journal of Indian Society Agric. Stat.* 23, 31-51.

Loius G. 1983: Successive sampling in large finite populations. *Annals of statistics* vol. 11, (2). 702-706.

Manish, T. and Shukla, D. 2008: *Efficient estimator in successive sampling using post stratification*.

Okafor, F.C. 1987: Comparison of estimators of population totals in two-stage successive sampling using auxiliary information. *Survey method.* 13(1), 109-121.

Okafor, F.C., Arnarb, R.1987: Some strategies of two-stage sampling for estimating population ratios over two occasions. *Aust.Journal stat.* 29(2) 128-142.

Okafor, F.C. 2001: *Treatment of non response in successive sampling*. University of Nigeria virtual library.

Okafor, F.C. 2002: *Sample survey theory with application*. Afro-Orbis publication, Nsukka.

Partterson, H.D. 1950: Sampling on successive occasions with partial replacement of units. *Journal of royal statistics society*. B12, 241-255.

Raj, D. 1965: On sampling over two occasions with probability proportional to size. *Ann.Math. Stat.* 36, 327-330.

Rao, J. N.K. and Graham, J.E. 1964: Rotation design for sampling on repeated occasions. *Journal American statistics Association*. 59, 492-509.

Rueda, M. Munoz, J.F., Gonzanez S. and Arcos, A. 2006: Estimating quartiles under sampling on two occasions with

arbitrary sample design. *Computational statistics and data analysis.* Vol. 51, issue 12. pp 6596 - 6613.

Rueda, M., Munoz, J.F and Arcos, A. 2007: Successive sampling in estimation quartile with p auxiliary variables. *Spring link journal (quality and quantity).*

Saito, G., Shinohara H. and Kazahaya K. (2002): Successive sampling of fumarolic gases at Satsuma-Iwojima and Kuju volcanoes, southwest Japan: Evaluation of short-term variations and recision of the gas sampling and analytical techniques. *Geochemical Journal*..36(1) 1-20.

Sen, A.R. (1971): Successive sampling with two auxiliary variables. *Sankya* B33, 371-378.

Sen, A.R .(1972): Successive sampling p-auxiliary variables. *Ann. Mathematical Statistics*.43, 203-204.

Shivtar, S. and Srivastava, A.K. (1973): Use of auxiliary information in two-stage successive sampling. *Journal Indian Society Agric. Statistics Association*. 25, 101-104.

Tikkiwal, B.D.(1967): Theory of multi-phase sampling from a finite population of successive occasions. *Rev. Inst. Statistics*. 35.(3) 247-263.

Test for the estimates of change with

different values of μ and ρ

μ	$\hat{V}\!\!\left(\hat{\Delta}_1 ight)$
0.0	0.0898
0.2	0.1101
0.4	0.1423
0.6	0.2011
0.8	0.3427
1.0	1.1584

Test for current estimates of variance with different

values of $ ho$			
$\stackrel{\scriptscriptstyle\wedge}{ ho}$	$\hat{V}\left(\hat{M}_{2} ight)$		
0.0	0.5792		
0.2	0.5734		
0.4	0.5550		
0.6	0.5213		
0.8	0.4634		
1.0	0.2896		



From t he above table minimum variance is achieved for current estimate when $\rho = 1.0$

^	1⁄4	1/3	1/2	1
$ ho \setminus \lambda$				
0.5	0.0938	0.1106	0.1250	0.080
0.6	0.1688	0.1990	0.2250	0.1440
0.7	0.3063	0.3611	0.4083	0.2613
0.8	0.6000	0.7075	0.8000	0.5120
0.85	0.9031	1.0650	1.2042	0.7707
0.9	1.5188	1.7909	2.025	1.2960
0.95	3.3844	3.9909	4.5125	2.8880

Relative efficiency or gain in % for different values of λ and ρ

The above table shows that substantial gains in precision are achieved by using the better estimator when $\rho = 1$ and the value of $\lambda = 0.5$ are unique and show maximum precision.

Test for estimates of average over time with :

(a) varying values of μ and fixed $\stackrel{\wedge}{\rho}$ = 1.9225



(b) Varying $\stackrel{\wedge}{\rho}$ for a fixed $\mu = 0.32$

p for a has			
$\hat{ ho}$	$\hat{V} \left(\stackrel{\wedge}{\Delta}_1 ight)$		
0.0	1.1584		
0.2	1.3065		
0.4	1.4378		
0.6	1.5549		
0.8	1.6602		
1.0	1.7552		

From the above tables, variance of estimate of sum gives the same value when $\rho = 0$ and $\mu = 1$. Thus independent samples should be taken for precision and improvement on each occasion.

Test for the current estimates with different values of $\stackrel{\wedge}{
ho}$

$\hat{\stackrel{_{\wedge}}{ ho}}$	$\hat{V}\left(\hat{M}_{2} ight)$
0.0	0.3929
0.2	0.3889
0.4	0.3765
0.6	0.3536
0.8	0.3143
1.0	0.1964

From the above table, minimum variance is achieved from current estimate of variance when $\rho = 1.0$

Test for the estimates of change with

(a) Varying μ and fixed $\hat{\rho} = 0.7385$



b) varying $\stackrel{\wedge}{\rho}$ fixed $\mu = 0.32$

v.	.34	
	$\hat{\stackrel{_{\wedge}}{ ho}}$	$\hat{V} \left(\hat{\Delta}_1 ight)$
Γ	0.0	0.7858
	0.2	0.6716
	0.4	0.5407
	0.6	0.3889
	0.8	0.2112
	1.0	0.000

From the above tables, variance of estimate of change gives the same value for $\stackrel{\wedge}{\rho} = 0$ and $\mu = 1$. Thus for small value of μ and a high value of $\stackrel{\wedge}{\rho}$, precision will be enhanced.

Test for the estimates of average over time with

(a) Varying values of μ and a fixed $\stackrel{\wedge}{\rho} = 0.7385$ (b) Varying $\stackrel{\wedge}{\rho}$ for a fixed $\mu = 0.32$,

μ	$\hat{V}\left(\hat{\Delta}\right)$	$\hat{\stackrel{\scriptscriptstyle\wedge}{ ho}}$	$\hat{V}\left(\hat{\Delta}\right)$
0.0	1.3660	0.0	0.7858
0.2	1.1902	0.2	0.8862
0.4	1.0545	0.4	0.9752
0.6	0.9466	0.6	1.0547
0.8	0.8587	0.8	1.1261
1.0	0.7858	1.0	1.1905

From the above table, variance of estimate of sum gives the same value when $\rho = 0$ and $\mu = 1$. Thus independent samples should be taken for precision and improvement on each occasion.