Comparison of Precision of Double Sampling and Sampling on Successive Occasions

Oshungade Isaac Olayiwola  
Statistics Department,  
University of Ilorin

Ajayi, Moses Adedapo  
Statistics Department, Federal Polytechnic, Ede, Osun-State, Nigeria  
Email: tkdapo@yahoo.com

Masopa, Adekunle Nurudeen  
Statistics Department, Federal Polytechnic, Ede, Osun-State, Nigeria

Abstract – The research was carried out on comparison of precision of double sampling and sampling on successive occasions for some predetermined values of the correlation coefficient. Double sampling for regression and ratio estimators were computed and compared with the result obtained from sampling on successive occasion. It was discovered that double sampling for regression has higher precision than double sampling for ratio even when the correlation coefficient is as low as 0.25. On the contrary double sampling for regression is not as good as sampling on successive occasions in terms of precision except for situations when percentage of the matched units is 90% regardless of the value of the correlation coefficient.

Keywords – Double Sampling, Comparison, Predetermined, Alternating and Occasion.

I. INTRODUCTION

Sampling is a versatile branch of Statistics which deals with the estimation of certain statistic with the use of appropriate estimators. The value of the sample statistic is in turn used to draw inferences about the population parameter. Thus, making sampling indispensable for virtually every researcher regardless of the specialization or/and areas of interest. However, there are many sampling methods with varying usages and feasibilities. The adoption of any sampling method or design for a given research depends on factors such as the population structure, how the variables changes over time and more essentially, the relative accuracy and precision of the sampling methods.

In cognizance of time-trend, samplers do prefer sampling methods involving the use of auxiliary information. It is conceived that the auxiliary variable will furnish us with some information about the actual variable of interest and thereby improving the precision of the sampling methods used. Raj\textsuperscript{10} mentioned that if there is one thing that distinguishes sampling theory from the general statistical theory, is the degree of emphasis laid on the use of auxiliary information for improving the precision of the estimates. More so, the researcher may be interested in estimating some statistics or parameter(s) on two occasions (say first and second occasions). Double sampling and its close associate sampling on successive occasions are typical examples of sampling methods that rely on auxiliary information but the latter gives room for estimating of parameter on different occasions. The procedure of selections in sampling on successive occasions is similar to the double sampling. Eckler\textsuperscript{2} asserted that sampling on two occasions with partial replacement of units can be regarded as Double Sampling. It is referred o as rotation sampling.

The two sampling methods involve taking initial sample (on one occasion) and second sample (on another occasion) are the two sampling methods make use of information contained in the first sample to form estimates on the second occasion and estimates of means and total depend on both occasions. The two sampling methods make use of auxiliary variable. For double sampling procedure x and for successive sampling x represent the auxiliary variable for the second occasions. In contrast, double sampling: units are selected from the initial sample while sampling on occasions requires selection of additional units (unmatched units) for estimation of means.

The main objective of this work is to compare the precision of the estimators of the two sampling methods. The implications of the correlation coefficients (\(\rho\)) and the percentage matched or subsample (\(\lambda\)) on the precision of the sampling methods were also investigated. For double sampling estimator, regression and ratio were considered.

II. LITERATURE REVIEW

Numerous researches have been carried out on Double Sampling and Sampling on two occasions. Raj\textsuperscript{10} compared regression and ratio estimators and established various conditions under which each of the methods is better than the other. The first attempt on sampling on successive occasions can be attributed to Jensen\textsuperscript{2}, in his work, he obtained two estimates: One was the sample mean based on the new sample units only and the second was a regression estimate based on the sample units observed on both occasions and an overall sample mean obtained on the first occasion.

Eckler\textsuperscript{2}, worked on different level rotation sampling where it was only established that for one level rotation sampling, only sampled values that have been drawn from the population of current time can be added to the sample pattern and at this situation, higher levels, both the earlier sample values and current values can be added. Eckler\textsuperscript{2} also developed the method of rotation sampling to obtain a minimum variance estimate of the population values (mean and total) by suitably constructing a linear function of sample values at different times. Patterson\textsuperscript{9} in his work, he derived a necessary and sufficient condition for a linear unbiased estimator to be a minimum variance estimator.

Yates\textsuperscript{12} extended Jensen’s\textsuperscript{2} result for the study of one character on two occasions to “\(h\)” (\(h\geq2\)) occasions under the restrictive conditions of a constant sample size and a fixed replacement fraction on each occasion. He also observed that the decrease in the correlation coefficient
between the same sampling units for the observed character on the different occasions followed a geometric progression. Singh\textsuperscript{11} studied the effect of the replacement policy on the variances when design is multi-stage and established that whether the sample is completely retained or replaced by a new sample, the variances of the estimates remains unchanged. He also developed unbiased linear estimates of means and variances on second and third occasions.

Raj\textsuperscript{10}, studied the work of Jessen\textsuperscript{4} and Cochran\textsuperscript{1} on the use of regression estimator in repetitive survey and derived an appropriate expression for the variance of sampling on two occasions. Pathak and Rao\textsuperscript{8}, demonstrated the inadmissibility of the customary estimators of the population total in sampling on two occasions and also proved that any convex loss function on the second occasions does not have greater precision than the first occasions. Improved estimators for both simple random sampling (SRS) and probability proportional to size (PPS) cases developed. Kulldu\textsuperscript{2}, showed that the estimators of the population total on the second occasion under simple random sampling without replacement have smaller large sample minimum variance than the one developed by Pathak and Rao\textsuperscript{8} for the same expected cost. Ghangurde and Rao\textsuperscript{7} established that the expected cost for two different special schemes are equal, assuming that the cost is proportional to the number of distinct units in the sample. Moreover, a necessary condition estimator based on a completely unmatched sample of the same size was developed.

Cochran\textsuperscript{1} asserted that where there is complete or no matching, the variance of sampling on two occasions reduces to variance of simple random sampling which will be greater than the minimum variance of sampling on two occasions. Furthermore, he established that the reductions in the variance are modest if the value of the correlation coefficient is less than 0.8.

Okafor\textsuperscript{6} compared some estimators of the population total in two stage successive sampling using auxiliary variable. He Okafor\textsuperscript{6} also considered the estimation of the population ratio on two successive occasions. Okafor and Arnab\textsuperscript{6} gave the estimator of the population total for two occasions under simple random sampling probability proportional to size estimation. It could be deduced sampling that the precision of the sampling methods relies on the number of matched or retained units as well as the correlation coefficients. Despite the closeness and adaptability of double sampling and sampling on successive occasions, comparative analysis on precision of the sampling methods has so far received little or no attention. This work is an attempt to uncover the precision preference between the sampling methods.

### III. METHODOLOGY

The validity and reliability of any research depend strongly on the methods used at various stages in the design of the survey, collection of data, analysis and interpretation of the result of the analysis. Since this work is centered on measure of precision which is a function of the variance(s), attention is focused on the unbiased variance estimators of the sampling methods.

The data is on students enrolment into Secondary Schools in Oyo State, Nigeria for two academic sessions (2008/2009 and 2011/2012) collected from the state Ministry of Education. The number of students in the selected schools in 2008/2009 sessions is taken as the auxiliary variable(x) and that of 2011/2012 is taken as the actual variable of interest (y). The number of students in the selected schools in 2008/2009 session was considered as (x) and the corresponding number of students for 2011/2012 session was considered as (y).

The selection procedure was based on (i) an initial sample of size n=40 is selected by simple random sampling without replacement (ii) an subsample n (20 < n <36) is taken from the initial sample and (iii) the number of students in selected schools were transcribed this was done to avoid complete matching and at least 50% matching. In the case of sampling on two occasions, the “n” subsample is referred to as the matched retained units “m”. This implies that n = m . However for estimation of means and total, sampling on two occasions requires taking n-m new units independent of the initial sample. The n-m units are referred to as the unmatched units on the second occasions.

Let \( \lambda \) and \( \theta \) be the percentage matched (sub sampled) and unmatched (new units) respectively. The following mathematical expressions hold

(i) \( \lambda = \frac{m}{n} \)

(ii) \( \theta = \frac{(n - m)}{n} \)

(iii) \( \lambda + \theta = 1 \)

Derivation of alternative estimators for the variances of the sampling methods

For adequate comparison, the adaptability of the estimators has to be duly looked into, this aspect discusses the transformation or derivation of the formulae for the variances of the sampling methods. The methods involves incorporation of some common features or parameters, such as \( \lambda \) and \( \rho_m \). The number of matched pairs (subsampled), correlation coefficient and the derivations are as follow:

**Double sampling Regression estimator**

Given that:

\[
\begin{align*}
\nu(y^1_{IR}) &= \left( \frac{N - M}{NM} \right) s^2_{my} + \left( \frac{n! - m}{n^m} \right) s^2_{my} (1 - \rho^2_m) \\
\nu(y^2_{IR}) &= \left( \frac{1}{m} - \frac{1}{N} \right) s^2_{my} + \left( \frac{n! - m}{n^m} \right) s^2_{my} (1 - \rho^2_m) \\
\nu(y^3_{IR}) &= \left( \frac{1}{m} - \frac{1}{N} \right) s^2_{my} + \left( \frac{1}{m} - \frac{1}{n} \right) s^2_{my} (1 - \rho^2_m) \\
\nu(y^4_{IR}) &= \frac{s^2_{my}}{m} + \left( \frac{1}{m} - \frac{1}{n} \right) s^2_{my} (1 - \rho^2_m)
\end{align*}
\]
\[ v(y'1r) = \frac{S^2_{my}}{m} + \frac{1}{m}(1-\lambda)S^2_{my}(1-\rho^2_m) \]
\[ v(y'1r) = \frac{S^2_{my}}{m} + \frac{1}{m}(1-\lambda)S^2_{my}(1-\rho^2_m) \]

The transformed regression estimator is:
\[ v(y'1r) = \frac{S^2_{my}}{m} \left\{ 1+(1-\lambda)S^2_{my}(1-\rho^2_m) \right\} \quad --(2) \]

**Ratio Estimator**

Given that:
\[ v(F^1_R) = \left( \frac{N-m}{Nm} \right) S^2_{my} + \left( \frac{n-m}{nm} \right) \left[ R^2_{m^2_{mx}} - 2R_m S_{my} \right] \quad --(3) \]
\[ v(F^1_R) = \left( \frac{1}{m} \right) S^2_{my} + \left( \frac{n-m}{nm} \right) \left[ R^2_{m^2_{mx}} - 2R_m S_{my} \right] \]
\[ v(F^1_R) = \left( \frac{1}{m} \right) S^2_{my} + \left( \frac{n-m}{nm} \right) \left[ R^2_{m^2_{mx}} - 2R_m S_{my} \right] \]
\[ v(F^1_R) = \left( \frac{s^2_{my}}{m} - \frac{1}{m} \left( 1 - \frac{m}{n} \right) \right) \left[ R^2_{m^2_{mx}} - 2R_m S_{my} \right] \]

The derivation of ratio estimator is:
\[ v(F^1_R) = \frac{1}{m} S^2_{my} + \left( \frac{1}{m} \right) \left( 1-\lambda \right) \left[ R^2_{m^2_{mx}} - 2R_m S_{my} \right] \quad --(4) \]

**Sampling on two occasions**

Given that,
\[ v(\mu) = \frac{(1-\rho^2_m)S_m}{(1-\rho^2_m)m} \quad --(5) \]
\[ s_m = \frac{(m-1)S^2_{my} + (m-1)S^2_{my}}{(m-1) + (m+1)} \]
\[ s_m = \frac{(m-1)S^2_{my} + (m-1)S^2_{my}}{2(m-1)} \]
\[ s_m = \frac{S^2_{my} + S^2_{my}}{2} \quad --(6) \]

The equations (5 and 6) derivation of sampling occasion becomes:
\[ v(\mu) = \frac{\left( 1-\rho^2_m(1-\lambda) \right) S^2_{mx} + S^2_{my}}{\left( 1-\rho^2_m(1-\lambda)^2 \right)} \quad --(7) \]

From the derivations of estimators above, it is obvious that the estimators of the sampling method have some cogent and precision determiner parameters or items in common. These derived variance estimators are used for the data analysis and conclusions shall be drawn in the light of the variance estimates of the different sample size and correlation levels.

**IV. DATA PRESENTATION AND DATA ANALYSIS**

This section presents the result of computations carried out by using the variance estimators of the sampling methods derived. The values are the calculated standard error of estimate at various (0.50 < \lambda > 0.90) with 0.05 interval between successive \lambda_s. Also, three fixed correlation coefficients \rho = 0.25, 0.50, 0.75 are chosen. This is to observe the behavior of the sampling methods for the cases of distributions considered as the values of ns and s changed.

The three correlation coefficients were chosen to represent the various possibilities of low (\rho = 0.25) intermediate (\rho = 0.50) and high (\rho = 0.75) correlations. This established that ratio regression estimators will be better than a simple random sampling (SRS) when \lambda > 0.5. In successive sampling with partial matching, the efficiency will depend on the correlation between the matched units. As stated in section 3.2 the values of n indicates the percentage of matched sample "m" For example, n=0.5 means half of the initial sample were retained or subsample. That is, if n=40 and n=0.5 it means 20 of the initial sample n were retained on the second occasion.

The application of the criterion for better precision stated earlier, the following indicators are used for easy appraisal and judgment of the precision hierarchy of the sampling method.

* * * Indicates (table II to IX below) that the sampling method have the highest precision for all correlation coefficients.

* * Indicates (table II to IX below) that the sampling method have a better precision at two correlation coefficients.

* Indicates (table II below) that the sampling method have a better precision at one correlation coefficient.

**Table I**

<table>
<thead>
<tr>
<th>\lambda=0.5</th>
<th>n'=40</th>
<th>n=m=36</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sampling method</td>
<td>\rho=0.25</td>
<td>\rho=0.50</td>
</tr>
<tr>
<td>Regression</td>
<td>187.5804</td>
<td>178.2730</td>
</tr>
<tr>
<td>Ratio</td>
<td>211.1662</td>
<td>189.1735</td>
</tr>
<tr>
<td>Sampling on two occasions</td>
<td>178.6813</td>
<td>174.0091</td>
</tr>
</tbody>
</table>

Copyright © 2014 IJISM, All right reserved
Discussions of findings

Based on the result in the tables (I to IX) presented above it can be observed that sampling on two occasions has the highest precision for \( \lambda = 0.50, 0.55 \) at \( \rho = 0.25, 0.50 \) but at the \( \rho = 0.75 \) for the same values of \( \lambda \), double sampling for regression has the highest precision while regression estimator dominated its double sampling counterpart, that is, ratio estimator. Finally, at \( \lambda = 0.90 \) at all values of \( \rho \), double sampling for regression has the highest precision, followed by sampling on two occasions while double sampling for ratio has the least precision. Though the result at this stage may have practical implications but it could as well be an outlier.

VII. CONCLUSION

It can be concluded that sampling on successive occasions is the best estimator for \( \lambda = 0.50, 0.55 \) at \( \rho = 0.25, 0.50 \) and other combinations of \( \rho \) and \( \lambda \) except for \( \lambda = 0.50, 0.55 \rho = 0.75 \) and \( \lambda = 0.90 \) for all value of \( \rho \) where double sampling for regression takes the lead and double sampling for ratio appears to be poorest estimator.

In the light of the results discussed above, it can be recommended that sampling on successive occasions should be employed when samplers or researchers have resources that can cater for between 50%-85% of the initial sample given that the sampling procedure is design in such a way that the final sample would be taken from a larger sample. Also, it can be recommended that in situations where samplers have sufficient resources that can cater for 90% of the initial sample to be taken at the second stage double sampling should be employed and for analytical reasons the regression form should be used to attain reasonable precision.
REFERENCES