# Comparison of precision of systematic sampling with some other probability samplings

## Habib Ahmed Elsayir

Dept. of Mathematics, Umm Al Qura University (K. S. A), Al Qunfudha University College

**Email address:**

Habibsayiroi@Yahoo.com

**Abstract:** This paper provides some discussion about the problem of precision of systematic and other related sampling methods. Comparison among these samples and the estimators are made using some empirical evidences from a finite population analysis.

**Keywords:** Bias, Systematic Sample, Precision, Random Sampling

## 1. Introduction

In this paper we try to discuss the precision in of some probability samplings with emphasis on systematic sampling. Precision is a measure of how close an estimator is expected to be to the true value of a parameter, which is usually expressed in terms of imprecision and related to the standard error of the estimator. Less precision is reflected by a larger standard error.*( Valerie J. Easto & John ,1997).* Sampling error and bias relate to precision and accuracy. A measurement is precise if it obtains similar results with repeated measurement (or repeated surveys).A measurement is accurate if it is close to the truth with repeated measurement (or repeated surveys). Because bias in survey, may lead to make inappropriate decisions about programmes based on invalid results, and lead to fail to provide needed services or waste resources on providing unneeded services. Bias may lead to grossly wrong conclusions, while having not quite enough precision may only decrease confidence in the survey results.

It is known that the precision of any estimate made from a sample depends both on the method by which the estimate is calculated from the sample data and on the plan of sampling .Therefore, the paper provides a general discussion on the related sampling methods with comparison on the basis of properties and precision.

Comparisons of Precision in systematic sampling and in other sampling methods is found in a wide range of literature. Some of these references are (*Sampford, 1962), (Cochran, William G.,1977), (Kirk M. Wolter, 1984)* and (*Kirk M. Wolter, 1985*), (*Vallian etal, 2000),* (*Brewer , 2002*).Recently,(*Crawford, I. M. , 1990),and( Crawford, I. M. , 1997)* used two major principles sample design to avoid bias in the selection procedure and to achieve the maximum precision for a given outlay of resources. *Rose(Rose AM, Grais RF, Coulombier D, Ritter H,2006)* compared the results of two different survey sampling techniques (cluster and systematic) where both survey methods gave similar results.*Megan(Megan Deitchler, Hedwig Deconinck and Gilles Bergeron ,2008)* gave a comparison of three sampling designs in an emergency setting ,and the paper considered the sampling precision of a systematic sampling method for estimating total number of nerve fibers exposed on cross section of a nerve trunk.

## 2. Comparing Systematic and other Sampling Methods

Systematic sampling is a method used when the population elements are arranged in a specific order from which a sample can be drawn in a systematic way rather than generate a simple random sample. Systematic sampling is especially applicable when the population to be studied is arranged in some order. In such cases it may be easier to draw random sample in a systematic way rather than generate a simple random sample. For example, a supermarket, which wishes to investigate customer attitudes, may interview each 3rd or 10th customer arriving to the shop. The same principle is applied to many phenomena. For example the temperatures in a city are

registered each hour in order to determine middle temperatures. It is straight forward that systematic sampling leads to a more representative survey than simple sampling. If, for example, the temperature is registered each day at 12 o'clock, the middle temperature would be wrongly estimated,*(Jorgen Lauridsen , 2005).*

To select a systematic sample of n elements from a population of N elements, we divide the N elements in the population to n groups of k elements and then randomly select the first element out of the first k elements in the population, and then we select every *k*-th unit afterwards until we have a sample of n elements. This type is called an every *k*-th systematic sample ( *Cochran, 1977).*If the units arranged in a circular way, and then take a random start from 1 to N and thereafter every *k*-th until n+1 units are selected, this is called circular systematic sampling ,*(Chakravarti etal , 1967).*If the units in the aggregate are first arranged in a linear order before selection, this is called linear systematic sampling, *(Chakravarti etal , 1967).*

In some situations, systematic samples tend to be more representative than could be simple random samples drawn from the same population.This depends upon the arrangement ,or frame from which the sample is drawn. If a serial correlation exists (when elements in the population tend to resemble other nearby elements more than they resemble those at greater distances),then the systematic sample will generally tend to be more representative than the simple random sample,*( Stehman S. V.,* 1992).

For instance, if the income of residents are investigated, and the sample is drawn by selecting every k-*th* house, from a complete map of the area, we will tend to get a more representative sample than would be obtained by a simple random sampling procedure. The systematic sample will be less subject to source of variability ,and for that reason, will tend to be more representative than the simple random sample and hence it is preferred in implementation because of its ease and design efficiency,*(Opsomer& Francisco , 2011). It can be* stated that the degree to which the systematic sample would be more representative than simple random sample depends upon the degree to which these tends to be a clustering of similar elements in the list or arrangement from which the sample was drawn. It is seemed that ,as the clustering becomes more obvious, the advantages of the systematic sample become greater .However, it is the simple random sample to which all the formulas introduced thus for apply. The systematic sample, on the other hand, is expected to be less representative than the simple random sample in situations where there is a cyclical or periodic movement of the data and the length of the period of the cycle tends to be equal to or close to k.

The method of systematic sampling has some advantages over other methods in general and over simple random sampling in particular. The main advantage of using systematic sampling over simple random sampling is its simplicity. *Joan (2009).*Some of these advantages are:

*(Opsomer &Francisco , 2011).*
1.   The actual selection of a systematic sample is easier and cheaper than that of a simple random sample ,and the population size need not be known exactly for systematic sampling.
2.   Any trend, or clumping of similar units, in the population ,will tend to be more actually represented in a systematic rather than in simple random sampling. Hence the systematic sample can be expected to give the more precise estimate.

### 2.1. Systematic &Stratified Random Sampling

Many populations are naturally divided into a number of non-overlapping subpopulations, denoted as strata or clusters. Generally the population is divided into K strata, each consisting of $N_k$ individuals. If the estimator of the mean in each stratum is unbiased, then the stratified estimator of the mean for the stratified sample is an unbiased estimator of the population mean, μ. In most cases stratified sampling leads to more efficient sampling than simple random sampling, so that higher precision(less variance) can be obtained in the estimates. *(Jorgen Lauridsen, 2005).*Precision can be calculated based on different allocations of the sample, i.e. choices of $n_1,$, $n_2,..._,n_k$ .

Viewed as cluster sample, systematic sample ordering the population and selecting each $(\frac{N}{n})'^{th}$ individual, can be viewed as a special case of stratified sampling: (*Opsomer & Francisco ,2011)*

Let stratum 1 be the first $\frac{N}{n}$ individuals, stratum 2 the next $\frac{N}{n}$ individuals etc. Choose a number $m \leq \frac{N}{n}$, and select individuals numbered the

$$m, m + \frac{N}{n}, m + 2\frac{N}{n}, ...$$

The systematic sample is spread more over the entire population than stratified sample, because in stratified sampling the samples in the strata are drawn separately .This adds precision in some cases. It can be noticed that cluster sampling may also be viewed as a special case of stratification, where stratified sampling is superior to random sampling because it reduces sampling error. Stratified random samplings generally have more statistical precision than simple random sampling.*( William M. G., 2006).* However, it can be stated that systematic random sampling is better than systematic sampling, but usually is not as good as simple random sampling, and is not much easier to implement,*( Philip B. Stark,* 2011).

## 3. The Variability

The precision of any estimate made from a sample depends both on the method by which the estimate is calculated from the sample data and on the plan of sampling. In the following analysis, let $y_{ij}$ denotes the *j*-th element of the *i*-th systematic sample ,so that *j*

=1,2,3,…,k .The mean of the $i$-th sample is denoted by $\bar{y}_i$ .The systematic sampling estimator of the population mean, μ .The variance of this mean estimator requires some assumptions as below:

1-When the population values are assumed to be in no particular order with respect to the variance of the estimator of the mean is the same as in the case of simple random sampling. (*Aczel, 1989).*

$$S^2\left(\overline{Y}_{sy}\right) = \frac{(N-n)}{Nn} S^2 \qquad (3.1)$$

2-When the mean is constant within each stratum of K elements but different from stratum to stratum, the estimated variance of the sample mean is:

$$S^2\left(\overline{Y}_{sy}\right) = \frac{\dfrac{N-n}{Nn}\displaystyle\sum_{i=1}^{n}\left(Y_i - Y_{i+k}\right)^2}{2(n-1)} \qquad (3.2)$$

3-When the population is assumed to be either increasing or decreasing linearly in the variable of interest, and when the sample size is large, the appropriate estimator of the variance of our estimator of the means given by: (*Aczel, 1989*)

$$S^2\left(\overline{Y}_{sy}\right) = \frac{\dfrac{N-n}{Nn}\displaystyle\sum_{i=1}^{n}\left(Y_i - 2Y_{i+k} + Y_{i+2k}\right)^2}{6(n-2)} \quad for\ 1 \le I^{/} \le n-2 \qquad (3.3)$$

According to (Cochran, 1977), the variance of the mean of a systematic sample is:

$$V\left(\overline{y}_{sy}\right) = \frac{N-1}{N}.S^2 - \frac{k(n-1)}{N}S_{wsy}^2 \qquad (3.4)$$

Where :

$$S_{wsy}^2 = \frac{1}{k(n-1)}\sum_{i=1}^{k}\sum_{j=1}^{n}\left(y_{ij} - \overline{Y}_i\right)^2$$

is the variance among units that lie within the same systematic sample. (*Cochran, 1977*) also showed that:

$$V\left(\overline{y}_{sy}\right) = \frac{S^2}{n}\frac{N-1}{N}\left[1 + (n-1)\rho_w\right] \qquad (3.5)$$

Where :

$\rho$ is the correlation coefficient between pairs of units that are in the same systematic sample. It is defined as:

$$\rho_w = \frac{E\left(y_{ij} - \overline{Y}\right)\left(y_{iu} - \overline{Y}\right)}{E\left(y_{ij} - \overline{Y}\right)^2}$$

Which gives:

$$\rho_w = \frac{2}{(n-1)(N-1)S^2}\sum_{i=1}^{k}\sum\sum_{j<u}\left(y_{ij} - \overline{Y}\right)\left(y_{iu} - \overline{Y}\right) \qquad (3.6)$$

And:

$$V\left(\overline{y}_{sy}\right) = S_{wst}^2 = \frac{1}{n(k-1)}\sum_{j=1}^{n}\sum_{i=1}^{k}\left(y_{ij} - \overline{Y}_j\right)^2$$

Where

$$\rho_{wst} = E\left(y_{ij} - \overline{Y}_j\right)\left(y_{iu} - \overline{Y}_u\right)$$

$$\rho_{wst} = \frac{2}{n(n-1)(k-1)}\sum_{i=1}^{k}\sum\sum_{j<u}\frac{\left(y_{ij} - \overline{Y}\right)\left(y_{iu} - \overline{Y}_u\right)}{S_{wst}^2}$$

The above quantity is the correlation between the deviations from the stratum means of pairs of items that are in the same systematic sample. Therefore a systematic sample has the same precision as the corresponding stratified sample, with one unit per stratum, if $\rho_w = 0$ . The variance of the mean from stratified sample is:

$$V\left(\overline{y}_{st}\right) = \left(\frac{N-n}{N}\right)\frac{S_{wst}^2}{n} \qquad (3.7)$$

and the variance $V\left(\overline{y}_{st}\right)$ is found directly from the systematic sample totals as:

$$V\left(\overline{y}_{st}\right) = V_{sy} = \frac{1}{k}\sum_{i=1}^{k}\left(\overline{y}_i - \overline{Y}\right)^2$$

$$= \frac{1}{n^2k}\sum_{i=1}^{k}\left(n\overline{y}_i - n\overline{Y}\right)^2 \qquad (3.8)$$

A great caution must be taken when considering variance estimation in systematic sampling. A danger in systematic sampling is that the characteristics being studied may have a certain pattern or periodicity in the list .On Census record sheets, for example, the first names on the sheets tend to be above average in income or rental value, for instance corner houses tends to have a higher rental value than houses in the middle of the block.

Used widely in surveys of finite population, this method picks up any obvious or hidden stratification in the population when used properly. It can be regarded as a random selection of one cluster; however, it is not always possible to obtain an unbiased or even consistent estimator of the design variance. Hence, biased estimators of variance must be sought if we are to estimate the precision of our investigation or survey estimators from the sample itself, (*Kirk*, 1984).

# 4. Empirical Results

To illustrate and verify results concerning the estimation of the Systematic sample compared with other samples, we consider data show the main annual food crops production of the Sudan for 25 years since 1970/1971. These crops are

sorghum, wheat and millet. Each of these is considered as a separate population, so we have population 1 see table (4-1), population 2, and population 3 for, sorghum, wheat and millet respectively, and the total population size N =25 for each population. For random and stratified sampling, analysis of variance of the population into" between rows" and "within rows" is desired. This is presented in tables (4-2),(4-4) and (4-6).

**Table (4-1).** *Data for 5 Systematic Samples with N=kn=25, Population (1):Sorghum Production:*

| strata | 1 | 2 | 3 | 4 | 5 | Strata means |
|--------|------|------|------|------|------|--------|
| Systematic sample numbers | | | | | | Strata |
| 1 | 1535 | 1592 | 1301 | 1691 | 1792 | 1582.2 |
| 2 | 2143 | 2606 | 2082 | 2353 | 1461 | 2129.0 |
| 3 | 2084 | 3335 | 1884 | 2006 | 1097 | 2081.2 |
| 4 | 3597 | 3277 | 1363 | 4425 | 1536 | 2839.6 |
| 5 | 1180 | 3581 | 4042 | 2386 | 3648 | 2967.4 |
| Totals | 10539 | 14391 | 10672 | 12861 | 9534 | 11599.4 |

Source: Data Analysis

**Table (4-2).** *Analysis of Variance*

| | d.f | s.s | m s |
|---|-----|-----|-----|
| Between rows(strata) | 4 | 6634830 | 1658707.5 |
| Within strata | 20 | 16134290 | $806714.5 = s^2_{wst}$ |
| Totals | 24 | 22769120 | $2465422 = s^2$ |

Source: Data Analysis

**Table (4-3).** *Data for 5 Systematic Samples with N=kn=25, Population (2):Wheat Production:*

| strata | 1 | 2 | 3 | 4 | 5 | Strata means |
|--------|------|------|------|------|------|--------|
| Systematic sample numbers | | | | | | Strata |
| 1 | 163 | 124 | 152 | 235 | 269 | 188.6 |
| 2 | 263 | 289 | 312 | 165 | 231 | 252.0 |
| 3 | 218 | 142 | 176 | 157 | 79 | 154.4 |
| 4 | 199 | 157 | 181 | 247 | 448 | 238.6 |
| 5 | 686 | 895 | 453 | 475 | 1436 | 591.4 |
| Totals | 1529 | 1607 | 1274 | 1279 | 1436 | 1425 |

Source: Data Analysis

**Table (4-4).** *Analysis of Variance*

| | d.f | s.s | m s |
|---|-----|-----|-----|
| Between rows(strata) | 4 | 617361.2 | 154340.3 |
| Within strata | 20 | 233282.8 | $11664.1 = s^2_{wst}$ |
| Totals | 24 | 850644.5 | $166004.4 = s^2$ |

Source: Data Analysis

**Table (4-5).** *Data for 5 systematic samples with N=kn=25, Population (3):Millet Production:*

| strata | 1 | 2 | 3 | 4 | 5 | Strata means |
|--------|------|------|------|------|------|--------|
| Systematic sample numbers | | | | | | Strata |
| 1 | 439 | 441 | 355 | 285 | 403 | 384.6 |
| 2 | 388 | 449 | 500 | 552 | 309 | 439.6 |
| 3 | 335 | 509 | 227 | 314 | 168 | 310.6 |
| 4 | 419 | 285 | 153 | 495 | 161 | 302.2 |
| 5 | 85 | 308 | 449 | 221 | 973 | 407.2 |
| Totals | 1664 | 1992 | 1684 | 1867 | 2014 | 1844.2 |

Source: Data Analysis

**Table (4-6).** *Analysis of Variance*

| | d.f | s.s | m s |
|---|-----|-----|-----|
| Between rows(strata) | 4 | 1443246.8 | 36011.7 |
| Within strata | 20 | 683703.2 | $34185.2 = s^2_{wst}$ |
| Totals | 24 | 2126950 | $394996.9 = s^2$ |

Source: Data Analysis

**Table (4-7).** *Summary of Results of Mean and Variance of Specific Sampling Methods.*

| | sorghum | wheat | millet |
|---|---------|-------|--------|
| | Pop 1 | Pop 2 | Pop 3 |
| Population mean N=25 | 2319.9 | 285 | 368.8 |
| $s^2$ | 2465422 | 16664.4 | 394996.9 |
| $V_{ran}$ | 473361 | 26560.7 | 63199.5 |
| $V(\bar{y}_{st})$ | 154889.2 | 1866.3 | 5469.3 |
| $V(\bar{y}_{sy})$ | 4968739.3 | 705.42 | 874.7 |

Source: Data Analysis

Now we are going to illustrate the analysis of variance of the population .The purpose for the analysis of variance is to test for significant differences between means. The within group variability is usually referred to as error variance. If the strata are homogeneous, the variability within-groups is expected to be lower than the variability for the population as a whole. From the table (4-2), it is seen that $V_{ran} > V(\bar{y}_{st})$. As from table (4-6), it is noticed that $V_{ran} > V(\bar{y}_{st}) > V(\bar{y}_{sy})$ .

It is seen that $V(\bar{y}_{st})$ is always less than $V_{ran}$ and $V(\bar{y}_{st})$ in population 1,2 and 3. But $V(\bar{y}_{sy})$ is less than $V_{ran}$ in two cases out of three as can be noted in population 2 and 3. This may be conducted to the distribution of the data. Results are presented in table(4-7)

The results show that both stratified random sampling and systematic sampling are much more effective than simple random sampling, but systematic sampling is less precise than stratified random sampling .Systematic sampling is more precise than simple random sampling because it picks up any hidden or obvious stratification in the population. It is clear that the variance of $\bar{y}$ is less for a systematic than for a simple random sampling since the list from which the systematic sample is drawn show a fairly consistent trend or grouping as the data represent production during series of time. This means that there may be a population for which systematic sampling is extremely precise and others for which it is less precise than simple random sampling. Therefore the performance of a systematic sampling in relation to that of stratified or simple random sampling is mainly dependent on the properties of the population. On the other hand, a systematic sample will have the same precision as the corresponding stratified random sample, with one unit per stratum, if $\rho_{wst = 0}$ which is the correlation between the deviation from the stratum means of pairs of items that are in the same systematic sample.

## 5. Discussion

The method of systematic sampling can be alternative to simple random sampling and specially preferred when the information required to construct a sampling is available in a list or any other organized form. Although systematic sampling is used for its convenience of drawing and execution, but its variance of estimator requires some assumptions about the underlying population order. Some of these assumptions are whether the population elements follow particular order, constant mean within strata, or the population is assumed to be either linearly increasing or decreasing in the variable of interest.

In the light of results, systematic sampling can safely be recommended in situations where the ordering of the population is essentially random or contains at most a mild stratification, for the effects of hidden periodicies tend to cancel out when systematic sample is drawn from each stratum.

As for a simple random sampling, it is a method which executed in a manner that every element of the population has equal probability of being included in the sample and a complete frame including units of population is needed. In addition, simple random sampling gives unbiased estimates of population means and variances estimates .Thus, when we do not select our sample randomly out of the population of interest, our sampling results may be biased. Hence, the necessity of simple random sampling arises.

It is seen that the performance of a systematic sampling in relation to that of stratified or simple random sampling is mainly dependent on the properties of the population. All the evidences show that both stratified random sampling and systematic sampling are much more effective than simple random sampling, but generally systematic sampling is less precise than stratified random sampling. The systematic sampling is approximately as precise as stratified random sampling with one unit per stratum. The difference between the two methods is that the systematic sample is spread more evenly over the entire population than a stratified random sample, because in the latter, the samples in stratified sampling the samples in the strata are drawn separately. This adds precision in some cases.

## 6. Conclusions

(1) When comparing simple random sampling, results show that the mean of a systematic sample can be expected to be more precise estimate of the population mean than the mean of a random sample of the same size. In addition, a larger sample size increases precision, and does not guarantee absence of bias, which may produce very incorrect results.

(2) The sampling variance of the mean of a systematic sample from a list can be expected to be less than that of the mean of a random sample, if there is a fairly consistent trend throughout the list or any appreciable grouping in the values of the population ,but this sampling variance can not be estimated without bias from a random systematic sample.

(3) Stratified sampling is known to produce weighted mean whose variability is less than that of arithmetic mean of a simple random sample of the population. This will help in applying random or systematic sampling in each of the stratum. This will also help in the reduction of sampling error since the representativeness of the samples are improved.

## References

[1] Aczel, Amir D(1989).Complete Business Statistics. Boston : Homewood.

[2] .Brewer, K.(2000) .Combined Survey Sampling Inference. London : Arnold.

[3] Chakravarti, I. M etal (1967).Handbook of methods of Applied Statistics. vol II :Planning of surveys and Experiments., New York: John Willey & Sons.

[4] Cochran, William G.(1977).Sampling Techniques, New York: John Willey & Sons.

[5] Crawford, I. M. (1990), Marketing Research, Centre and Network for Agricultural Marketing Training in Eastern and Southern Africa, Harare, pp 36-48.

[6] Crawford, I. M. (1997).Marketing Research and information systems, Food and Agriculture Organization(FAO) of the United Nations ,Rome.

[7] Joan ,Joseph Castillo (2009).Systematic Sampling. Experiment Resources, in htt://www.experiment-resources.com.

[8] Jørgen Lauridsen(2005). Module 5: Systematic and stratified sampling (1:4) http://statmaster.sdu.dk/courses/st118/module05

[9] Kirk M. Wolter (1985).Introduction to Variance Estimation. Second Edition. Series of Statistics for Social and Behavioral Sciences. Springer.

[10] Kirk M. Wolter, (1984).Variance Estimation for Systematic Sampling . Statistics for Social and Behavioral Sciences, 298-353. Springer. com.

[11] Megan Deitchler, Hedwig Deconinck and Gilles Bergeron(2008) .Precision, time, and cost: a comparison of three sampling designs in an emergency setting ,Statistics Glossary, V11. Emerging Themes in Epidemiology 2008, 5:6 doi:10.1186/1742-7622-5-6 © 2008 Deitchler et al; licensee BioMed Central Ltd. (http://creativecommons.org/licenses/by/2.0

[12] Ministry of Agriculture(1996) .Time Series of Cereals and Oil Crops 1970-71-1995-96 No.3 Sudan.

[13] Opsomer, J. D. &Francisco Fernandez, M.(2011).Model-based Nonparametric variance Estimation for Systematic Sampling. in: www.stat.colostate.edu/~jopsomer/.../Nonpara_model_variance.

[14] Philip B. Stark(2011). SticiGui Department of Statistics University of California, Berkeley website: http://www.stat.berkeley.edu/~stark/SticiGui/index.htm.

[15] Rose AM, Grais RF, Coulombier D, Ritter H(2006). A comparison of cluster and systematic sampling methods for measuring crude mortality. Bull World Health Organ 2006, 84(4):290-6Bull World Health Organ. 2006 Apr; 84(4):290-6. Epub 2006 Apr 13.

[16] Sampford (1962).An Introduction to Sampling Theory with Applications to Agriculture. Edinburgh: Oliver and Boyd.

[17] Stehman S. V.(1992). Comparison of systematic and random sampling for estimating the accuracy of maps generated from remote American Society for Photogrammetry and Remote Sensing, Bethesda, MD. ly sensed data. Photogrammetric. 1992, vol. 58, n 9, pp. 1343-1350 .

[18] Valerie J. Easton& John H. McColl Statistics Glossary Vol 1 webpage ihttp://www.stats.gla.ac.uk/steps/glossary/sampling.html.

[19] Vallian, R. Dorfman, A. and Royall, R(2000).Finite population Sampling and Inference: John Wiley and Sons, New York.

[20] William M. G. (2006).Probability sampling, Research Methods Knowledge Base in Social research methods.net/samppr.

## Biography



**Dr. habib Ahmed Elsayir** –associate prof in statistics, received his Ph. D degree in Statistics from Omdurman Islamic University, Sudan in 2001. He was appointed manager of Omdurman Islamic University Branch at Al Daein 2002-2005. He was appointed the Head Dept. of Mathematics (2010-2012), Alqunfudha University college, University of ummal Qura, Saudi Arabia.