https://doi.org/10.56093/JISAS.V80I1.5
Available online at https://isas.live/newsite
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 80(1) 2026 43–48

Applications of Neutrosophic Statistics in Analysing Systematic Random Sample Survey Data

Pranesh Kumar

University of Northern British Columbia, BC, Canada

Received 19 January 2026; Revised 25 March 2026; Accepted 25 March 2026

SUMMARY

Systematic random sampling method has been extensively used to select samples when populations are defined across time and space and when sampling frame is not determined. This method is preferred for its ease of implementation and simple procedure, low budget considerations, convenience of sample selection and also, even when sampling frame is not available. In this paper, we aim to present different types of systematic sampling methods for the estimation of population mean/total. For illustration purpose, we have considered the real data examples from the practical applications and have analyzed them using neutrosophic logic, which takes into account the indeterminate or imprecise or missing data information.

Keywords: Neutrosophic logic and math; Linear systematic sample; Circular systematic sample; Survey data analysis.

MSC: 62D05, 94D05

This paper is dedicated to the special issue of The Indian Society of Agricultural Statistics in the memory of Professor Dr. Prem Narain, who had made a significant contribution in the field of Agricultural Statistics. At IASRI, Dr. Narain has motivated me and channeled my professional journey from 1970 until 1988 in various capacities, first being a teacher, and afterwards, as the director. I have failed to find suitable words to express my gratitude to him and just pray for him.

1. INTRODUCTION TO SYSTEMATIC SAMPLING

The first reference on systematic sampling is noted in the paper by Madow and Madow (1944). Systematic sampling is a simpler and cost-effective probability sample selection method, which is generally applied in cases where interest is in sampling the large natural areas at a regular interval. For selecting a systematic sample, the method consists of choosing a random number (random start), which constitutes the first sample unit. The remaining sample units are selected according to a predetermined pattern. Samples may be evenly distributed over space or time or ordered. Systematic samples can be drawn in several ways. For example, a framework or network may be used to determine the sampling points. Sampling is done at the chosen or nearby approachable point. If sampling is done across a transect line, sampling points could be identified in some pre-determined order. To identify transect lines, we draw the vertical lines running from the top to bottom and divide the map from west to east or draw the horizontal lines running left to right and divide the map from north to south. Cochran (1946) has done some foundational work on the systematic sampling and carried out some numerical studies for the cases in which the correlogram is linear and exponential. Yates (1948) has discussed the results of an investigation into one-dimensional systematic sampling in which the sampling of sequences of quantitative values was undertaken using the sampling points equally spaced along the sequence. Iachan (1982) has presented the critical systematic sampling review and has also developed systematic sampling with emphasis on asymptotic and optimality results under the framework

E-mail address: pranesh.kumar@unbc.ca

of super-population models. Lovasi, Fink, Mooney and Link (2017) have studied design-based inference for systematic sampling, variance estimation and replication methods when health relevant context is indexed. The model-based perspective, emphasis is on the processes producing outcome variation, and observed data are used to make inference about that process while in the design-based perspective, emphasis is on the inference when the population is well defined and finite, as well as commonly observed in complex survey samples. These two perspectives play important role in deciding whether clustering must be accounted for analytically and how to select cluster definitions. Mostafa and Ahmad (2016) have considered remainder linear systematic sample with multiple random starts. Dumelle, Higham, Ver Hoef, Olsen, and Lisa (2022) have compared the design-based and model-based approaches to statistical inference in a finite population spatial context.

2. ADVANTAGES AND DISADVANTAGES OF SYSTEMATIC SAMPLING

The systematic sampling is one of the most useful sampling methods because it is easier to select a systematic sample, particularly, when the sample units are selected in the field or forest or in water sources. In addition, systematic sampling can be more efficient than the simple random sampling when explicit or implicit stratification is available in the sampling frame. Bias in systematic sampling may arise if the starting point (random start) is chosen non-randomly. Unknown periodicity in the population unit list affects the variance but not the property of unbiasedness of the estimate when the start is random. An inherent pattern in the population, coinciding with the sampling interval, may cause biasness in systematic sampling. Importantly, it may, however, be noted that the systematic sampling has the bias and limitation of estimating the variance estimate from a single systematic sample. Thus, wherever applicable, one should use methods of interpenetrating sub-samples or replications to obtain valid error estimates and reporting these choices transparently for the sake of clarity. Systematic sampling suffers from two main limitations. Firstly, the systematic sampling assumes that the sampling interval k is an integer. If population size $N \neq nk$, systematic sampling has a variable sample size, which depends on the random start. Consequently, the sample mean will become a biased estimator for the population mean as referred by Kish (1965). Cochran (1977) considered systematic sampling as a cluster sampling when only one cluster is randomly selected, each of size n units. Thus, Madow and Madow (1944) noted that a single systematic sample alone could not be used to estimate the design variance of the sample mean, i.e., sampling variance.

3. FUZZY AND NEUTROSOPHIC LOGIC

Fuzzy logic was first proposed by Zadeh (1965) as an expansion of the classical and multi valued logics. The basic idea of fuzzy logic is from probability that an event can have a probability between 1 (certain to happen) and 0 (certain not to happen). According to fuzzy logic, phrases such as likely, not likely, probably, unlikely, most, few, usually and nearly all are much more common than the absolute (precise) values of true and false. He further studied similarity relations and fuzzy orderings (1971).

Smarandache (2014) has proposed the neutrosophic logic, which is an extension/combination of the fuzzy logic, intuitionistic logic, para consistent logic, and the three-valued logics that use an indeterminate value. According to neutrosophic logic, every logical variable x is described by an ordered triple (t,i,f), where t is the degree of truth, f is the degree of false and i is the level of indeterminacy. For the sake of consistency with the classical and fuzzy logics and with probability, there is a special case where $t+i+f=1$. However, referring to the intuitionistic logic, which means incomplete information on a variable, proposition or event, we have $t+i+f<1$. Analogically, in para consistent logic, which means contradictory sources of information about a same logical variable, proposition, or event one has $t+i+f>1$.

4. LINEAR SYSTEMATIC SAMPLING (LSS)

Table 1 presents the symbols and notations used in what follows now. Table 2 provides the arrangement of population units for linear systematic sampling. To select a sample of size n from a population of size N such that N is multiple of n, i.e., $N=nk$, we first arrange the nk population units in k columns, each of size k. First, a random number (or, random start)

$r(1 \le r \le k)$ is selected, where k is termed as the sampling interval. Then, with the random start r, every k-th unit is selected. Thus, the desired systematic sample contains the population units with serial numbers r, $r+k$, ..., $r+(n-1)k$. In case of $N \neq nk$, we choose $k$as the integer nearest to N/n. In linear systematic sampling, we divide the N population units into k groups, each group of exactly n units in a systematic manner, and then we select one of these groups with probability $1/k$ that constitutes the desired systematic sample. Since, each population units occurs only once in one of the k samples, the inclusion probability of every population unit remains the same. As compared to simple random sampling, systematic sampling results in more efficient estimates of population mean/total, when populations units follow linear trend and where study variable values are linearly related with the serial numbers of units in the population. Systematic sampling is also found efficient in populations where population units have autocorrelation (1996).

Denoting the sampled i-th observation by $[y_{iL}, y_{iU}]$, the estimate of total $[Y_L, Y_U] = \sum_{i=1}^{N} [y_{iL}, y_{iU}]$ is

$$\left[\hat{Y}_L, \hat{Y}_U\right]_{lsys} = N \sum_{i=1}^{N} [y_{iL}, y_{iU}] / n \,, \tag{1}$$

estimate of variance is

$$v\left[\hat{Y}_L, \hat{Y}_U\right]_{lsys} = \frac{N(N-n)\sum_{i=1}^{n-1}[(y_{i+1} - y_i)_L^2, (y_{i+1} - y_i)_U^2]}{2n(n-1)}. \tag{2}$$

Confidence interval is commonly preferred over the point estimate, which results in a single value. The 95% confidence interval for the population total is given by

$$= \text{Estimate} \pm 2(\text{Standard error of the estimate})$$

$$= \left[\hat{Y}_L, \hat{Y}_U\right]_{lsys} \pm 2\sqrt{v\left[\hat{Y}_L, \hat{Y}_U\right]_{lsys}}. \tag{3}$$

Now, we consider the data in the neutrosophic number form in the example below from the e-Book "Elements of Survey Sampling", Exercise 6.6, page 175 [10].

Example 1. It is desired to estimate the average per day rent for single occupancy rooms in well-known hotels of an Indian state. In all, there are 192 such hotels listed in a book entitled "A Guide to Visitors". The investigator selected a l-in-8 sample of hotels and rang up the managers of sampled hotels. The information on rent (in rupees) was obtained. Data is expressed in neutrosophic format and is given in Table 3. Purpose is to estimate the average per day rent along with the confidence limits.

Thus, the population size is N=192 and the sample size is n=24. Hence, the sample interval is $k=N/n$=8. Let the random number r selected between 1 and $k=8$ be 6.

Estimated value of the average rent per day using estimator in equation (1) is

$$\left[\hat{\bar{Y}}_L, \hat{\bar{Y}}_U\right]_{lsys} = \frac{[2560, 2685]}{24} = [106.67, 111.88] \text{Rupees},$$

and the estimate of variance estimator using equation (2) is

$$v\left[\hat{\bar{Y}}_L, \hat{\bar{Y}}_U\right]_{lsys} = \frac{(192-24) * [16600, 17375]}{192 * 24 * (24-1)} = [568.15, 594.68].$$

The 95% confidence interval for the average rent per day from equation (3) is given by

$$[106.67, 111.88] \pm 2 * \sqrt{[568.15, 594.68]} = [154.34, 160.65] \text{Rupees}.$$

Thus, the estimate of the average rent per day from the single systematic sample of 24 hotels from the list of 192 hotels is $[106.67, 111.88]$Rupees. The 95% confidence interval, obtained above, indicates that the average rent per day is likely to be in the interval $[154.34, 160.65]$ with the probability 0.95.

5. CIRCULAR SYSTEMATIC SAMPLING (CSS)

In circular systematic sampling, for selecting a sample of size n from the population of size N, first $N(N=nk \text{ or } N \neq nk)$ population units are arranged around a circle (1996). Then, we select a random start r $(1 \le r \le N)$ and we choose the unit corresponding to the random start as the first unit to be included in the sample. Thereafter, we select every k-th unit arranged on the circle until a sample of desired size n units is

chosen. Thus, population units corresponding to the serial numbers $r+jk$ (if $r+jk \le N$) and $r+jk-N$ (*if* $r+jk > N$), $j = 0, 1, 2, \ldots,$ (n-l), will be selected for the sample (1996).

For estimating population total $Y = [Y_L, Y_U] = \sum_{i=1}^{N} [y_{iL}, y_{iU}]$, the estimator is

$$\left[\hat{Y}_L, \hat{Y}_U\right]_{csys} = N \sum_{i=1}^{N} [y_{iL}, y_{iU}] / n, \quad (4)$$

and the estimate of variance

$$v\left[\hat{Y}_L, \hat{Y}_U\right]_{csys} = \frac{(N-n)\sum_{i=1}^{n}\left[[y_{iL}, y_{iU}] - \left[\hat{Y}_L, \hat{Y}_U\right]_{csys}\right]^2}{nN(n-1)}, \quad (5)$$

The 95% confidence interval for population total is given by

$$\left[\hat{Y}_L, \hat{Y}_U\right]_{csys} \pm 2\sqrt{v\left[\hat{Y}_L, \hat{Y}_U\right]_{csys}}. \quad (6)$$

Example 2. For the purpose of estimating the total fish catch at the end of a particular day, 162 fishing boats had gone to sea from the coast (Pages 152-153, Example 6.4 [10]). 15 boats were selected using circular systematic sampling and the total fish catch from them were recorded. Data expressed in neutrosophic format and is given in Table 4. Obtain the estimate of total catch of fish and also, the 95% confidence limits for the total fish catch.

The estimated total catch of fish at the end of the day from the sample of 15 boats in quintals using equation (4) is

$$\left[\hat{Y}_L, \hat{Y}_U\right]_{csys} = 162 * \frac{[111.6, 114.8]}{15} = [1205.28, 1237.68] \text{ quintals},$$

and the estimate of variance using equation (5) is

$$v\left[\hat{Y}_L, \hat{Y}_U\right]_{csys} = \frac{162 * (162-15) * [69.48, 66.03]}{2 * 15 * (15-1)} = [3743.9, 3939.5].$$

The 95% confidence interval for total fish catch $[Y_L, Y_U] = \sum_{i=1}^{N} [y_{iL}, y_{iU}]$ from equation (6) is given by

$$[1205.28, 1237.68] \pm 2 * \sqrt{[3743.9, 3939.5]} = [-1079.7, 1115.3] \text{ quintals}.$$

Thus, the estimate of total fish catch from a single sample is $[1205.28, 1237.68]$ quintals. The confidence limits, obtained above, indicate that the total fish catch is likely to fall in the interval $[-1079.7, 1115.3]$ quintals.

6. ESTIMATING POPULATION TOTAL/ MEAN USING INTERPENETRATING SUBSAMPLES

Interpenetrating subsamples are used as another method for estimating population total/mean of required size n. In this method, we select two or more (say m) systematic subsamples of same size with independent random starts [1]. Let m interpenetrating subsamples, each of size n/m, are to be chosen from the N population units and $N/n = k$. For selecting the required m samples with linear systematic sampling, we select m random starts with replacement or without replacement from *1* to mk (assumed to be integer). Thus, the subsample corresponding to a particular random start will contain the population units with serial numbers at intervals of mk. If mk is not an integer, we can use circular systematic sampling for the selection of the subsamples.

Let $\left[\bar{y}_{1L}, \bar{y}_{1M}\right], \ldots, \left[\bar{y}_{mL}, \bar{y}_{mM}\right]$ be the population mean estimators based on m subsamples, each of size n/m. Then, the unbiased estimator of population mean

$$\widehat{\left[\bar{Y}_L, \bar{Y}_M\right]}_{sy} = \sum_{i=1}^{m} \left[\bar{y}_{iL}, \bar{y}_{iM}\right] / m, \quad (7)$$

estimate of variance is

$$v\widehat{\left[\bar{Y}_L, \bar{Y}_M\right]}_{sy} = \sum_{i=1}^{m} \left(\left[\bar{y}_{iL}, \bar{y}_{iM}\right] - \widehat{\left[\bar{Y}_L, \bar{Y}_M\right]}_{sy}\right)^2 / m(m-1), \quad (8)$$

and the 95% confidence interval for population total is given by

$$\widehat{\left[\bar{Y}_L, \bar{Y}_M\right]}_{sy} \pm 2\sqrt{v\widehat{\left[\bar{Y}_L, \bar{Y}_M\right]}_{sy}}. \quad (9)$$

Example 3. A dairy research institute wishes to estimate the total milk yield of a buffalo with reference to its breeding program (Page 154, Example 6.5, [10]). In Table 5, total lactation period was taken as 300 days. It was decided to select 3 systematic subsamples each of size 10 days, so as to arrive at a total sample size of 30 days.

Since, $N = 300$, $n = 30$, and $m = 3$, we have $k = 300/30 = 10$. We select without replacement (WOR)

samples so that no subsample is repeated. Since we decided to choose three subsamples, we select three random starts from *1* to *mk = 30*. Let random starts be 1, 12, and 26. Corresponding to these random starts, the selected days and corresponding milk yields in liters recorded in neutrosophic data are given in Table 5. Subsample means are also given at the end of the table.

The estimate of total milk yield (in liters) obtained by using (7) is

$$\widehat{N\left[\overline{Y}_L, \overline{Y}_M\right]}_{sy} = N\sum_{i=1}^{m} \frac{\left[\overline{y}_{iL}, \overline{y}_{iM}\right]}{m}$$

$$= \frac{300}{3}\left[\left[10.0, 10.14\right] + \left[9.68, 9.86\right] + \left[9.62, 9.73\right]\right]$$

$$= \left[2929.10, 2973.30\right] \text{liters},$$

estimate of variance is

$$N^2 v\widehat{\left[\overline{Y}_L, \overline{Y}_M\right]}_{sy} = N^2 \sum_{i=1}^{m} \frac{\left(\left[\overline{y}_{iL}, \overline{y}_{iM}\right] - \widehat{\left[\overline{Y}_L, \overline{Y}_M\right]}_{sy}\right)^2}{m(m-1)}$$

$$= \frac{30^2\left(\left[0.0071, 0.1409\right] + \left[0.0552, 0.0101\right] + \left[0.0847 + 0.0011\right]\right)}{3(3-1)}$$

$$= \left[1983.99, 2053.02\right],$$

and the 95% confidence interval for population total is given by

$$\widehat{\left[\overline{Y}_L, \overline{Y}_M\right]}_{sy} \pm 2\sqrt{v\widehat{\left[\overline{Y}_L, \overline{Y}_M\right]}_{sy}} = \left[1176.93, 6897.07\right] \text{liters}.$$

Thus, the total milk yield, based on the sample of the 300 observations, would most probably with 95% confidence be observed between the interval [1176.93, 6897.07] liters.

7. CONCLUDING REMARKS

Systematic random sampling is commonly used sampling technique to select samples from the populations across time and space because of the convenient sample selection procedure and low cost. We have presented foundational scientific work of systematic sampling and its development, their advantages and disadvantages, and the population mean/total estimation methods. Further, since collected data may be imprecise or indeterminate, Smarandache has proposed the neutrosophic logic, which is an extension/combination of the fuzzy logic, intuitionistic logic, para consistent logic, and the three-valued logics that use an indeterminate value. For the ready reference, we have included a general description of neutrosophic logic. For illustration purpose, we have included the real data examples and their analyses using neutrosophic logic. For the easy access of the variance, confidence intervals formulas and the neutrosophic assumptions, Appendix 1 is added. Manual calculations using MS Excel were carried out to do the neutrosophic calculations like addition, subtraction, multiplication and division.

As a possible future work, we plan to continue with investigating the analysis of systematic samples since the neutrosophic arithmetic used to construct confidence intervals sometimes yields impossible bounds like negative values for the intrinsically non-negative quantities. This indicates that the variance/CI construction under neutrosophic intervals does not respect positivity constraints and likely over-propagates uncertainty. However, we think that it may be so because neutrosophic value has three components true, false, and indeterminate. The indeterminate component may cause large and inadmissible values. Further, we plan to study a rigorous framework for interval/uncertainty propagation (e.g., constraint-aware intervals, set-valued CIs based on optimization, or simulation-based coverage studies under assumed neutrosophic sets), together with proofs or empirical evidence showing that intervals remain feasible and of reasonable width. Another line of work will be to study the feasibility of the intervals produced or on when the approach is preferable to standard design-based inference with imputation/measurement-error models.

ACKNOWLEDGEMENTS

Author is thankful to the referee for the review of the paper and comments, which helped in revising the manuscript. He also thanks University of Northern British Columbia, Prince George, Canada for providing the facilities where the research was carried out.

REFERENCES

Cochran, W.G. (1946). Relative accuracy of systematic and stratified random samples for a certain class of populations. Annals of Mathematical Statistics **17**, 164-77.

Cochran, W.G. (1977). Sampling techniques. New York, NY: John Wiley & Sons.

Dumelle, M, Higham, M, Ver Hoef, J.M., Olsen, A.R. and Lisa Madsen, L. (2022). A comparison of design-based and model-based approaches for finite population spatial sampling and inference, Methods in Ecology and Evolution, https://doi.org/10.1111/2041-210X.13919

Iachan, R. (1982). Systematic sampling: A critical review. International Statistical Review **50**, 293-303.

Iachan, R. (1983). Asymptotic theory of systematic sampling. Annals of Mathematical Statistics **11**, 959-969.

Kish, L. (1965). Survey sampling. New York, NY: John Wiley & Sons.

Lovasi, G., Fink, D.S., Mooney, S. and Link, B.G. (2017). Model-based and design-based inference goals frame how to account for neighborhood clustering in studies of health in overlapping context types. SSM Population Health, 3, 600-608. https://doi.org/10.1016/j.ssmph.2017.07.005

Madow, L.H., and Madow, W.G.(1944). On the theory of systematic sampling. Annals of Mathematical Statistics, **15**, 1-24.

Mostafa, S.A., and Ahmad, I.A. (2016). Remainder linear systematic sample with multiple randomstarts. Journal of Statistical Theory and Practice, **10**(4), 824-51.

Singh, R. and Mangat, N.S. (1996). Survey Sampling. Springer Science+Business Media, Dordrecht.

Smarandache, F. (2014). Introduction to Neutrosophic Statistics. Sitech & Education Publishing.

Yates, F. (1948). Systematic sampling. Philosophical Transactions of the Royal Society of London, **241**, 345-77.

Zadeh, L.A. (1965). Fuzzy sets. Information and Control, vol. 8 (1965), pp. 338-353.

Zadeh, L.A. (1971). Similarity relations and fuzzy orderings. Information Sciences, **3**, 177-200.

Table 1. Symbols and Notations

N	Population size	T	True
N	Sample size	I	Indeterminate
K	Sampling interval	F	False
R	Random start	$y_i=[y_{iL}, y_{iU}]$	y-value of the i-th observation
$V(\overline{y}_{sys})$	*Variance of sample mean estimator*	*Lsys*	Linear systematic sample
$v\left[\hat{Y}_L, \hat{Y}_U\right]$	*Variance estimator of sample mean*	*Css*	Circular systematic sample

Table 2. Arrangement of $N=nk$ Population units for Linear Systematic Sampling

1	*2*	...	*r*	...	*K*
k+1	*k+2*	...	*k+r*	...	*2k*
2k+1	*2k+2*	...	*2k+r*	...	*3k*
...	...	...	...	...	...
...	...	...	...	...	...
...	...	...	...	...	...
(n-1)k+1	*(n-1)k+2*	...	*(n-1)k+r*	...	*Nk*

Table 3. Average hotel room rent (in rupees) from the sampled 24 hotels

Hotel	Rent (Rs)	Hotel	Rent (Rs)	Hotel	Rent (Rs)
1	[95, 100]	9	[90, 95]	17	[120, 125]
2	[120, 125]	10	[105, 110]	18	[80, 85]
3	[125, 130]	11	[120, 125]	19	[90, 95]
4	[115, 120]	12	[80, 85]	20	[100, 105]
5	[110, 115]	13	[70, 75]	21	[125, 130]
6	[75, 80]	14	[120, 125]	22	[90, 95]
7	[125, 130]	15	[125, 130]	23	[130, 135]
8	[120, 125]	16	[100, 105]	24	[130, 140]

Table 4. Catch of fish (in quintals) from the sampled 15 boats

Boat Serial Number	y Fish Catch	Boat Serial Number	y Fish Catch	Boat Serial Number	y Fish Catch
73	**[5.4, 5.7]**	128	[9.1, 9.3]	21	[8.3, 8.5]
84	[8.0, 8.3]	139	[6.6, 6.7]	32	[10.7, 10.9]
95	[5.9, 6.2]	150	[7.2, 7.4]	43	[6.9, 7.0]
106	[9.6, 9.8]	161	[5.7, 5.9]	54	[5.4, 5.6]
117	[8.3, 8.7]	10	[6.8, 6.9]	65	[7.7, 7.9]

Table 5. Milk Yield (in liters) for 30 selected days

Sub-sample	I	Sub-sample	II	Sub-sample	III
Day	Milk Yield (liters)	Day	Milk Yield (Liters)	Day	Milk Yield (Liters
1	[8.00, 8.20]	12	[9.20, 9.40]	26	[11.00, 11.20]
31	[12.00, 12.00]	42	[13.48, 13.52]	56	[14.70, 14.70]
61	[15.00, 15.40]	72	[14.20, 14.80]	86	[14.50, 14.70]
91	[14.00, 14.20]	102	[14.20, 14.40]	116	[12.80, 12.80]
121	[11.25, 11.27]	132	[9.80, 9.80]	146	[10.60, 10.70]
151	[10.10, 10.12]	162	[10.00, 10.20]	176	[10.50, 10.60]
181	[9.80, 9.80]	192	[8.40, 8.80]	206	[8.30, 8.30]
211	[8.70, 8.80]	222	[8.38, 8.42]	236	[7.40, 7.60]
241	[7.00, 7.50]	152	[6.00, 6.20]	266	[4.30, 4.30]
271	[4.10, 4.10]	282	[3.10, 3.10]	296	[2.10, 2.40]
Mean	[10.0, 10.14]		[9.68, 9.86]		[9.62, 9.73]