Estimating the Ratio of a Crisp Variable and a Neutrosophic Variable

Carlos N. Bouza-Herrera[1,*], **Mir Subzar**[2]

[1] Universidad de La Habana La Habana , Cuba
[2] Jammu and Kashmir Institute of mathematical sciences (JKIMS), J&K, India
Email: subzarstat@gmail.com; bouza@matcom.uh.cu

Abstract
The estimation of the ratio of two means is studied within the neutrosophic theory framework. The variable of interest Y is measured in a sample of units and the auxiliary variable X is obtainable for all units using records or predictions. They are correlated and the sample is selected using simple random sampling. The indeterminacy of the auxiliary variable is considered and is modeled as a neutrosophic variable. The bias and variance of the proposed estimator are derived.

Keywords Ratio of Means; Neutrosophic Variable; Bias; Mean Squared Errors;Taylor Series.

1. Introduction

Atanassov (1999), considered how the soft set theory constitutes a general mathematical tool for modeling uncertainty and impreciseness. This approach overcomes the need of parameterizing, as in the theories of probability, fuzzy sets and rough sets . The neutrosophic theory is based on a new set conceptualization. The roots of it may be found in Smarandache (2002, 2003). Neutrosophic set is a generalization of the so-called intuitionistic fuzzy sets. This theory is being used in different areas of knowledge. See recent examples as Ajay, (2020), Crespo Berti (2020) in modeling real life problems; Hatip (2020) and Saqlain et al. (2020) who developed extensions of it.

Neutrosophic theory may be used for characterizing the indeterminacy of a variable. In real life problems the parameters of a population are unknown, and statistical inferential procedures usually replace them by an adequate estimation. Therefore, the unknowledge of the value of a parameter is overcome by determining approximate values of it. The statistician knows that the obtained values are imprecise. The inaccuracy is measured by using some formulas that provide a confidence level or an estimation error of the estimation. See recent discussions in ratio

DOI: 10.5281/zenodo.4030309

estimation in Bouza and Alomari (2019), where solutions are derived by using auxiliary information for ranking the pre-selected units, and Subzar et al. (2019) who derived robust alternative ratio estimators.

Neutrosophic theory provides a new framework for dealing with impreciseness. It is well known that using a Neutrosophic point of view statistical concepts and methods may expanded, see Smarandache (2013, 2014), Schweizer (2020), Cacuango et al. (2020). for example. This theory deals not with crisp values of the variables, but with set values.

Neutrosophic statistical data analysis issues have been discussed in different papers, and alternative tools have been derived on the domain of neutrosophic sets. In general, Neutrosophic Statistics may be considered as an alternative to classical statistics. Decision makers may be concerned with applications where to deal with Neutrosophy in the sample is needed. See for example the papers of Hanafy, et al. (2012) who studied neutrosophic correlation in this context, Patro and Smarandache (2016) and Alhabib et al. (2018), who dealt with distributions. This paper is highly motivated by the contributions of Aslam (2018, 2019b), Aslam, et al (2020b), where sampling based tools for control charts neutrosophic random variables were developed as well as the applications of Neutrosophic statistical tools reported in Aslam (2019a) and Aslam et al. (2020a).

A basic structure in neutrosophic theory is the set

$$A_N = \{\langle \xi, T_A(\xi), I_A(\xi), F_A(\xi)\rangle | \xi \in \chi\};\ (\xi) \neq \phi \qquad (1.1)$$

$$T_A(\xi) = degree\ of\ membership \qquad (1.2)$$

$$I_A(\xi) = degree\ of\ indeterminacy, F_A(\xi) = degree\ of\ non\ membership \qquad (1.3)$$

$$\upsilon_A(\xi)\chi \longrightarrow [0,1], \upsilon = T, I, F; \upsilon \in\ ]\ {}^{-}0, 1^{+}[\qquad (1.4)$$

From a practical point of view the interval [0,1] is used due to the difficulties that arise from using $]\ {}^{-}0, 1^{+}[$.

For developing estimation methods in the context of this study neutrosophic numerical measures play a key role. The following results are particularly important.

Take $\forall\gamma \in R, Z \neq 0, Z \neq -W$ and holds that $\gamma Z + \gamma W I_Z = \gamma(Z + W I_Z)$. Then

1. $\forall\gamma \in R, Z \neq 0, Z \neq -W$ holds that $\frac{\gamma I_Z}{Z+WI_Z} = \frac{\gamma}{Z+W} I_Z$
2. $\forall Z \neq 0, Z \neq -W$ holds that $\frac{I_Z}{Z+WI_Z} = (Z + WI_Z)^{-1} I_Z$
3. $\forall Q \neq 0$ holds that $\frac{Z+WI_Z}{Q} = \frac{Z}{Q} + \frac{WI_Z}{Q}$
4. $\forall Z \neq 0, Z \neq -W$ holds that $\frac{Q}{Z+WI_Z} = \frac{Q}{Z} + \frac{QW}{Z(Z+W)} I_Z$

DOI: 10.5281/zenodo.4030309

See Smarandache (2013, 2014) for details.

The statistical procedures, based on equations and formulas, may be generalized, by using the framework provided by neutrosophic theory. Substituting an estimate of a parameter θ by a set value, say $\hat{\theta}_N$, the estimate determines a neighborhood of $\hat{\theta}$, which includes the point estimate $\hat{\theta}$. The impreciseness associated to the point estimator $\hat{\theta}$ is included in the neutrosophic representation of the estimate.

An open-minded statistician will agree with this modeling. In real life, a sample is selected from the population but when the statistics are computed the situation has changed. That is, the true value of the variable is different from the value measured and used in the computations. Consider for example a study of the persons infected with Covid-19. The researcher selects a sample for developing PSR-tests and the degree of infection of the selected persons is measured, say Z. When the laboratory ends processing the collection of tests, the real value of Z in the patients may be very different. Hence, the decision making rule should consider that the value Z is imprecise. Determining a neighborhood of the computed estimate would be more realistic. Referring only to randomness, for modeling uncertainty, is myopic in many cases. Developing particular neutrosophic statistical methods should allow dealing with randomness and indeterminacy at the same time.

Let us provide a theoretical frame for sampling. Take a finite population of items

$$U = \{u_1, \dots, u_M\} \tag{1.5}$$

A random sample is selected from U using a sample design d. The probabilistic model is characterized by $\{U, d\}, The$ sample space S is an adequate algebra. The sampler fixes to use as sampling design a certain probability measure

$$d(s) = probability\ of\ observing\ a\ subset\ s \subseteq U, \sum_{s \in S} d(s) = 1 \tag{1.6}$$

Taking Y as the variable of interest and u_i as an item of U the evaluation is represented by $Y(u_i)=Y_i$. A random sample s of size m is selected by using the sample design d. If $u_i \in s$ then ,Y_i is a random variable. In classical statistics Y is evaluated in the units selected and an estimate (a statistic) is computed. The unknown parameter θ is estimated using an estimator

$$\hat{\theta} = \hat{\theta}\big(Y(s)\big), Y(s) = \{Y_1, \dots, Y_m | \forall i = 1, \dots, m; u_i \in s \ \} \tag{1.7}$$

The behavior of a statistic is evaluated considering how it behaves in the long run. That is, how is its performance when analyzing a long sequence of samples $s_1, .., s_P, P \to \infty$.

DOI: 10.5281/zenodo.4030309

Considering neutrosophic statistics an extension of the classical statistics may be developed using the similar principles. When the statistician considers that the data is known only approximately , he/she would not be confident in the inferences generated using crisp numbers. For example if a question is sensitive the respondent will tend to falsify the true value of Y, if carrying the stigma. Then the data is expected to be indeterminate. This situation is present in different neutrosophic studies as in the problems analyzed in Cacuango et al. (2020).

Neutrosophic statistics framework admits that the information provided by the data is not crisp, but ambiguous, vague, imprecise and/or incomplete. When the indeterminacy is zero the analysis made using a neutrosophic point of view would coincide with the results derived by classical statistics. The practitioner, using neutrosophic statistical methods, would interpret and organize the data taking into account the existence of these indeterminacies for obtaining some clues on the underlying patterns.

Some basic ideas on the estimation within a neutrosophic theory framework are developed in the next section. Section 3 is concerned with the development of some aspects of estimation theory in a Neutrosophic context. Section 4 develops a theory on ratio estimation. Numerical experiments are discussed in the Section 5.

2. Estimation in a Neutrosophic context

The proposals of Smarandache (2014 , 2016) discussed how common statistic equations and formulas, due to the data indeterminacy of some of the involved variables, may be better replaced by considering that they take values in a fixed set. The usual notation is to replace the variable crisp Z by its neutrosophic counterpart Z_N. N identifies that the variable is "neutrosophic". The impreciseness on the true value of Z is modeled by considering not a value but a set including it. In the applications of statistics the decision makers frequently deal with imprecise data. A convenient model seems to be considering $Z_{iN} = Z_i + A_i I_Z$ instead of Z_i The statistician measures Z_i but has the feeling that it is imprecise. It is subject to a basic error, which belongs to an interval I_Z, and is "tuned" by A_i for each "i". For example, a promoter obtains in the web a value of the index of achievement of a singer, say Z_i , but doubts that it is correct, as changes in the public preferences are constant. Hence, the change in the singers indexes moves in I_Z =(-2,5 2,5) but for a particular singer the decision maker considers that it may be 3,5 times larger. Then, is recorded Z_i+3,5I_Z. Note that now the decision maker is able to implement decision rules where impreciseness is modeled.

Consider a sample of size m, the observations determine the sequence $\{(Z_i + A_i I_Z), i = 1, \ldots m\}$. From Smarandache (2014 , 2016), is easily deduced that the sample mean is

$$\bar{Z}_{N(m)} = \bar{Z}_m + \bar{A}_m I_z = \frac{1}{m}\left(\sum_{i=1}^{m} Z_i + \sum_{i=1}^{m} A_i I_Z\right) \tag{2.1}$$

The deviation of each observation is

DOI: 10.5281/zenodo.4030309

$$D_{N(m)i} = (Z_i + A_i I_Z) - (\bar{Z}_m + \bar{A}_m I_z) \quad (2.2)$$

and its square is given by

$$D^2_{N(m)i} = \left[D_{N(m)i} = (Z_i + A_i I_Z) - (\bar{Z}_m + \bar{A}_m I_z)\right]^2 \quad (2.3)$$

Therefore, the sample variance is

$$s^2_{Z_{N(m)}} = \frac{1}{m}\sum_{i=1}^{m}(Z_i - \bar{Z}_m)^2 + \frac{1}{m}\sum_{i=1}^{m}[2(Z_i - \bar{Z}_m)(A_i - \bar{A}_m) - (A_i - \bar{A}_m)^2]I_Z \quad (2.4)$$

because $(Z_i - \bar{Z}_m + A_i I_Z - \bar{A}_m I_z)^2 = (Z_i - \bar{Z}_m)^2 + [2(Z_i - \bar{Z}_m)(A_i - \bar{A}_m) + (A_i - \bar{A}_m)^2]I_Z$

(2.1) and (2.4) estimate the neutrosophic parameters

$$\bar{Z}_N = \bar{Z}_M + \bar{A}_M I_z = \frac{1}{M}\left(\sum_{i=1}^{M} Z_i + \sum_{i=1}^{M} A_i I_Z\right) \quad (2.5)$$

$$\sigma^2_{ZNM} = \frac{1}{M}\sum_{i=1}^{M}(Z_i - \bar{Z}_M)^2 + \frac{1}{M}\sum_{i=1}^{M}[2(Z_i - \bar{Z}_M)(A_i - \bar{A}_M) - (A_i - \bar{A}_M)^2]I_Z \quad (2.6)$$

The interest of the decision maker is concerned with estimating a ratio of two variables. One of them is measured and the other is obtained from records. For example, we may develop an inquiry and obtain an achievement index for singers of the disk company. The records obtained in the web provide information on the downloads of a song but they are considered imprecise. Then the decision maker has a crisp variable, obtained in the inquiry, and a neutrosophic one obtained from the web.

In the sequel the sample design considered is Simple Random Sample With Replacement (SRSWR). The paper is concerned with the estimation of the ratio of a crisp variable and a neutrosophic one. The approximated bias and variance of the proposal are derived in the next section.

3. Some considerations on sampling

The importance of sampling experiments in different fields of applied sciences is one of the most important achievements of statistical inferences. The model considers the existence of a finite population U= $\{u_1, \dots, u_M\}$. The units are well identifiable. The researcher is interested in estimating a function of a variable Y. It is well defined for each individual of the finite population U. It is useful assuming that the experimenter have the knowledge of an auxiliary variable X for any individual of U $\{X(u_i) = X_i, i = 1, \dots, M\}$ but Y is unknown. The sampler may use the known values of X in the inference process. Under some mild conditions we may develop models, which yield more accurate results including the information provided by X. It is common that we should deal with indeterminacies in the values of X and is needed to determine how the model is affected. Hence, as the recorded values of X are imprecise an alternative is to use the neutrosophic number $X+AI_X$.

DOI: 10.5281/zenodo.4030309

The sample experiments are described as usual. For the population U there is a sample space S. The sampler selects a sample design d(s). It assigns a probability to each element of the sample space and allows determining the probability of selecting a certain unit of U by computing

$$P(u_i) = \sum_{\{s\in S|i\in s\}} d(s) \quad (3.1)$$

The variable of interest Y is measured in each selected unit, it is a random variable y_i that provides a result $Y(u_i)$. The sampler looks in the collection of recorded values of X, $X_1, \dots, X_M$ and obtains the corresponding random value x_i. Due to the existence of some indeterminacy is considered that it is the neutrosophic random variable

$$X_i + I_{X_i}, I_{X_i} = A_i I \ . \quad (3.2)$$

Then, in the study should be considered the existence of indeterminacy in the records and acknowledging this fact to work within a neutrosophic framework. Hence, taking an individual u_i of the population $u_i \rightsquigarrow \left(Y_i, X_i + I_{X_i}\right), I_X = A_i I$ is obtainable from it. As the unit is selected with probability $P(u_i)$ the gathered information is random. Using the sample design d the expectation of the result of the random experiment may be determined. Performing simple operations with neutrosophic numbers for an observation it is

$$E_d\left(y_i, x_i + I_{x_i}\right) = \sum_{j=1}^{M}\left(Y_j, X_j + A_j I_X \ \right) P\left(u_j\right) = \left(\sum_{j=1}^{M} Y_j\, P\left(u_j\right), \sum_{j=1}^{M} X_j\, P\left(u_j\right) + \left(\sum_{j=1}^{M} A_j\, P\left(u_j\right)\right)\ I_X\right) = \left(\mu_{1Y}, \mu_{1X} + \mu_{1A} I_X\right). \quad (3.2)$$

In the rest of the paper, without losing in generality, is considered that A_j=A for any j=1,…,M.

The sampling design to be considered in this paper is Simple Random Sampling (SRS) Without replacement. It is defined, see Singh (2003), as

$$d(s) = \begin{cases} \frac{1}{\binom{M}{m}}\ if\ \|s\| = m \\ 0\ otherwise \end{cases} \quad (3.3)$$

In that case

$$\forall i = 1, \dots, M, P(u_i) = \frac{m}{M}. \quad (3.4)$$

This result also holds if the selection is made with replacement. When M is sufficiently large the difference between selecting with or without replacement is negligible, in terms of the inferential processes, when

$$\frac{M-m}{M} \cong 1. \quad (3.5)$$

DOI: 10.5281/zenodo.4030309

A SRS sample is selected and are observed the realizations $\{y_i, x_i + A_i I_X \ , \forall i = 1, \dots, m\}$. Under SRS $P(u_i)$ is constant and using (3.2) the expectation is

$$E_d(y_i, x_i + I_{x_i}) = (\bar{Y},\ \bar{X} + \bar{A} I_X), \bar{Z} = \frac{1}{M}\sum_{j=1}^{M} Z_j, Z = Y, X \qquad (3.6)$$

because, for this design

$$\mu_{1Z} = \bar{Z} = \frac{1}{M}\sum_{j=1}^{M} Z_i;\ \ Z = X, Y, A \qquad (3.7)$$

Using (2.6) is easily derived that the variance of $x_i + I_{x_i}$is given by

$$V_d(x_i + I_{x_i}) = \sigma^2_{X_M} = \frac{1}{M}\sum_{i=1}^{M}(X_i - \bar{X}_M)^2 + \frac{1}{M}\sum_{i=1}^{M}[2(X_i - \bar{X}_M)(A_i - \bar{A}_M) - (A_i - \bar{A}_M)^2] I_Z =$$

$$= \sigma^2_X + \tau_{X,A} I_X \qquad (3.8)$$

Y is a crisp variable and its variance is

$$V_d(y_i) = \sigma^2_{Y_M} = \frac{1}{M}\sum_{i=1}^{M}(Y_i - \bar{Y}_M)^2 \qquad (3.9)$$

4. A ratio estimator

Frequently statistical research must deal with the estimation of a ratio or use it for deriving an estimator of the mean or the total of a variable of interest Y. A concomitant , or auxiliary, variable X, correlated with Y, is known. The population ratio of them is $R = \frac{\bar{Y}}{\bar{X}}$. Consider that a SRS sample is drawn from the population. A naive estimator is $\hat{R} = \frac{\bar{y}}{\bar{x}}$, where $\bar{y}$ and $\bar{x}$ are the sample means of Y and X. The auxiliary variable is obtained, commonly, from records or predictions, which usually are outdated and/or subject to impreciseness. The impreciseness may be modeled adequately in the context of Neutrosophic Theory. Consider the neutrosophic number $+AI_X, I_X \in (a, b)$. The sampler may model the imprecise knowledge on X by determining that for every individual of the population a measurement error interval. For example, if the decision maker considers that the percent of tax under-report is between 0% and 20%, is fixing that $I_X \in (0, 0{,}2X)$.

The variable Y is measured by direct interviewing the individuals of the population. It is a non-neutrosophic number. In the previous example, R may be

- The rate of the mean of the preferences of the public for a song with respect to the monthly mean of downloads.
- The rate of the reported taxes with respect to the previous occasion payments.
- The ratio of the fuel consumption reports of a transport enterprise in two consecutive months.
- The ratio of monthly incomes of employees versus the use of their credit card.

DOI: 10.5281/zenodo.4030309

The classical sampling theory assumes the non-existence of impreciseness in X. The sampler, being uncertain on the preciseness of the values of X, fixes (x_0, x_1) and works with the neutrosophic value $X_N = X_i + A_i I_{X_i}; I_{X_i} \in (x_0, x_1), i = 1, \dots, M.$ Take the crisp ratio of the means of two variables Y and X

$$R = \frac{\frac{1}{N}\sum_{i\in U} Y_i}{\frac{1}{N}\sum_{i\in U} X_i} = \frac{\bar{Y}}{\bar{X}} \tag{4.1}$$

X is known but is expected that it may change when the data is processed for computing. As the values of X come from imprecise recorded data, the decision maker fixes a conservative rule using $Q = \bar{Y}, Z = \bar{X}, W = \bar{X}$. In the context of neutrosophic R is denoted as

$$R = \frac{Q}{Z + W I_Z} = \frac{Q}{Z} + \frac{QW}{Z(Z+W)} I_Z = \frac{\bar{Y}}{\bar{X}}; Q = \bar{Y}, Z = \bar{X}, W = 0, I_Z \in [0,1]. \tag{4.2}$$

The alternative neutrosophic population ratio is

$$R_N = \frac{\bar{Y}}{\bar{X} + \bar{X} I_{\bar{X}}} = \frac{\bar{Y}}{\bar{X}} + \frac{\bar{Y}}{\bar{X}^2 + \bar{X}} I_{\bar{X}} = R + R^* I_{\bar{X}} \tag{4.3}$$

The operator $\frac{\bar{Y}}{\bar{X}^2 + \bar{X}} = R^*$ is used for modeling the indetermination present in the unknowledge of the true value of $\bar{X}$. Values of $I_{\bar{X}}$ close to zero model the decision maker`s confidence that $\bar{X}$ is the true expectation of X. In many occasions is adequate using

$$I_{\bar{X}} = \frac{1}{N}\sum_{i=1}^{N} I_{X_i} = I_X, I_X \in (x_0, x_1). \tag{4.4}$$

The decision maker usually fixes x_0=0.

Once a sample is selected the sample mean of X_N is computed

$$\bar{x}_N = \bar{x} + \bar{A}_m I_X;\ \bar{x} = \frac{1}{m}\sum_{i=1}^{n} x_i \tag{4.5}$$

as well as

$$\bar{y} = \frac{1}{m}\sum_{i=1}^{n} y_i \tag{4.6}$$

The need of estimating a ratio of two variables in this context suggests using $\bar{A}_m = \bar{x}$. Then, the neutrosophic estimator is

$$\hat{R}_N = \frac{\bar{y}}{\bar{x}_N} = \frac{\bar{y}}{\bar{x}} + \frac{\bar{y}}{\bar{x}^2 + \bar{x}} I_X = r + r^* I_X \tag{4.7}$$

DOI: 10.5281/zenodo.4030309

Take the first term and analyze

$$r - R = \frac{\bar{y}}{\bar{x}} - \frac{\bar{Y}}{\bar{X}} \tag{4.8}$$

Consider that $\Delta_{\bar{z}} = \frac{\bar{z}-\bar{Z}}{\bar{Z}}, Z = X, Y$. As $|\Delta_{\bar{x}} < 1|$ with $lim_{t\to\infty} \Delta_{\bar{x}}^{t} = \infty$ is valid developing $\Delta_{\bar{x}}$in the Taylor Series

$$r \cong R(1 + \Delta_{\bar{y}})(1 + \Delta_{\bar{x}})^{-1} \cong R\left[1 + \Delta_{\bar{y}} - \Delta_{\bar{x}} + \Delta_{\bar{x}}^{2} - \Delta_{\bar{y}}\Delta_{\bar{x}} + O(\Delta_{\bar{x}})\right]$$

The expected values of the summands are $E_d(\Delta_{\bar{z}}) = 0, z = x, y$ $E_d(\Delta_{\bar{x}}^2) = \frac{\sigma_X^2}{n\bar{X}^2}, E_d(\Delta_{\bar{y}}\Delta_{\bar{x}}) = \frac{\sigma_{XY}}{n\bar{X}\bar{Y}}$ and is derived that

$$E_d(r - R) \cong \frac{\sigma_x^2}{n\bar{X}^2} - \frac{\sigma_{xy}}{n\bar{X}\bar{Y}} \tag{4.9}$$

For the second term

$$(r^* - R^*)I_X \cong \left(\Delta_{\bar{y}} + \Delta_{d\bar{x}}^2 - \Delta_{\bar{y}}\Delta_{d\bar{x}} + O(\Delta_{d\bar{x}})\right) I_X \tag{4.10}$$

where

$$\Delta_{D\bar{x}} = \frac{(\bar{x}^2+\bar{x})-(\bar{X}^2+\bar{X})}{\bar{X}^2+\bar{X}} \tag{4.11}$$

The approximation is valid assuming that $\Delta_{D\bar{x}} \cong 0$ for a sufficiently large sample size. The first term of the expected value is $E_d(\Delta_{\bar{y}}) = 0$. Note that

$$\Delta_{D\bar{x}}^2 = \left(\frac{(\bar{x}^2-\bar{X}^2)+(\bar{x}-\bar{X})}{\bar{X}^2+\bar{X}}\right)^2 \tag{4.12}$$

Its design expectation is

$$E_d(\Delta_{D\bar{x}}^2) = \frac{1}{(\bar{X}^2+\bar{X})^2}\left[(\mu_{\bar{x}4} - \bar{X}^4) + (\mu_{\bar{x}3} - \bar{X}^3) + \frac{\sigma_X^2}{m}\left(1 - 2(\bar{X}^2 + \bar{X})\right)\right] \tag{4.13}$$

On the other hand

$$\Delta_{\bar{y}}\, \Delta_{D\bar{x}} = \left(\frac{\bar{y}-\bar{Y}}{\bar{Y}}\right)\left(\frac{(\bar{x}^2+x)-(\bar{X}^2+\bar{X})}{\bar{X}^2+\bar{X}}\right) \tag{4.14}$$

and

$$E_d(\Delta_{\bar{y}}\, \Delta_{d\bar{x}}) \cong \frac{1}{m}\left(\frac{1}{\bar{Y}}\right)\left(\frac{1}{\bar{X}^2+\bar{X}}\right)\left(m\sigma_{\bar{y}\bar{x}^2} + \sigma_{YX} - \bar{Y}\sigma_X^2\right) \tag{4.15}$$

Hence we may state that

DOI: 10.5281/zenodo.4030309

$$Bias\ (\hat{R}_N) \cong \left[\frac{\sigma_{\bar{X}}^2}{m\bar{X}^2} - \frac{\sigma_{XY}}{m\bar{X}\bar{Y}}\right] + \frac{1}{(\bar{X}^2+\bar{X})}\left[\left[(\mu_{\bar{x}4} - \bar{X}^4) + (\mu_{\bar{x}3} - \bar{X}^3) + \frac{\sigma_{\bar{X}}^2}{m}\left(1 - 2(\bar{X}^2 + \bar{X})\right)\right] + \frac{1}{m}\left(\frac{1}{\bar{Y}}\right)\left(\frac{1}{\bar{X}^2+\bar{X}}\right)\left(m\sigma_{\bar{y}\bar{x}^2} + \sigma_{\bar{y}\bar{x}} - \bar{Y}\sigma_{\bar{X}}^2\right)\right] I_X \tag{4.16}$$

Then, we have proved the following result.

Proposition 1. The expectation of the estimator $\hat{R}_N = \frac{\bar{y}}{\bar{x}_N} = \frac{\bar{y}}{\bar{x}} + \frac{\bar{y}}{\bar{x}^2+\bar{x}} I_X = r + r^* I_X$ of the neutrosophic ratio $R_N = \frac{\bar{Y}}{\bar{X}} + \frac{\bar{Y}}{\bar{X}^2+\bar{X}} I_{\bar{X}} = R + R^* I_{\bar{X}}$, when the sample is selected using simple random sampling, is approximately

$$E_d(\hat{R}_N) \cong R_N + \left[\frac{\sigma_{\bar{X}}^2}{\bar{X}^2} - \frac{\sigma_{XY}}{m\bar{X}\bar{Y}}\right] + \frac{1}{(\bar{X}^2+\bar{X})}\left[\left[(\mu_{\bar{x}4} - \bar{X}^4) + (\mu_{\bar{x}3} - \bar{X}^3) + \frac{\sigma_{\bar{X}}^2}{m}\left(1 - 2(\bar{X}^2 + \bar{X})\right)\right] + \frac{1}{m}\left(\frac{1}{\bar{Y}}\right)\left(\frac{1}{\bar{X}^2+\bar{X}}\right)\left(\sigma_{\bar{y}\bar{x}^2} + \sigma_{\bar{y}\bar{x}} - \bar{Y}\sigma_{\bar{X}}^2\right)\right] I_X;\ \ \mu_{\bar{x}t} = E_d(\bar{X}^t), \tag{4.17}$$

when the sample size is sufficiently large for accepting that both $\Delta_{\bar{x}^2}$ and $\Delta_{\bar{x}}$ are approximately equal to zero for a large sample size m.□

For deriving the Mean Squared Error (MSE) consider

$$\left(\hat{R}_N - R_N\right)^2 = \left(\frac{\bar{y}}{\bar{x}} + \frac{\bar{y}}{\bar{x}^2+\bar{x}} I_X - \frac{\bar{Y}}{\bar{X}} - \frac{\bar{Y}}{\bar{X}^2+\bar{X}} I_{\bar{X}}\right)^2 = (r - R + r^* I_X - R_N^* I_X)^2 = (r - R)^2 + (r^* I_X - R_N^* I_X)^2 + 2((r - R)(r^* I_X - R_N^* I_X)) \tag{4.18}$$

Developing the first term in taylor series is obtained

$$(r - R)^2 \cong \Delta_{\bar{y}}^2 + \Delta_{\bar{x}^2}^2 - 2\Delta_{\bar{y}}\Delta_{\bar{x}} \tag{4.19}$$

Hence

$$E_d(r - R)^2 \cong \frac{\sigma_{\bar{Y}}^2}{m\bar{Y}^2} + \frac{\sigma_{\bar{X}}^2}{m\bar{X}^2} - 2\frac{\sigma_{YX}}{m\bar{Y}\ \bar{X}} \tag{4.20}$$

The second term is

$$(r^* - R^*)^2 I_X \cong \left(\Delta_{\bar{y}}^2 + \Delta_{D\bar{x}}^2 - 2\Delta_{D\bar{y}}\Delta_{d\bar{x}} + O\left(\Delta_{D\bar{x}}\right)\right) I_X \tag{4.21}$$

and its expectation is given by

$$E_d((r^* - R^*)^2 I_X) \cong \left[\frac{\sigma_{\bar{Y}}^2}{m\bar{Y}^2} + \frac{1}{(\bar{X}^2+\bar{X})^2}\left[(\mu_{\bar{x}4} - \bar{X}^4) + (\mu_{\bar{x}3} - \bar{X}^3) + \frac{\sigma_{\bar{X}}^2}{m}\left(1 - 2(\bar{X}^2 + \bar{X})\right)\right] - 2\frac{1}{m}\left(\frac{1}{\bar{Y}}\right)\left(\frac{1}{\bar{X}^2+\bar{X}}\right)\left(\sigma_{\bar{y}\bar{x}^2} + \sigma_{YX} - \bar{Y}\sigma_{\bar{X}}^2\right)\right] I_X \tag{4.22}$$

Developing the third term is obtained

$$((r-R)(r^{*}I_X - R_N^{*}I_X)) \cong (\Delta_{\bar{y}} - \Delta_{\bar{y}} + \Delta_{\bar{x}}^2 - \Delta_{\bar{y}}\Delta_{\bar{x}})(\Delta_{\bar{y}} + \Delta_{D\bar{x}}^2 - \Delta_{D\bar{y}}\Delta_{D\bar{x}}\ )I_X \cong \left(\Delta_{\bar{y}}^2 - \Delta_{\bar{y}}\Delta_{\bar{x}}\ \right)I_X \qquad (4.23)$$

because only the terms of order t≤2 in the Taylor Series are considered as significant. Therefore

$$E_d((r-R)(r^{*}I_X - R_N^{*}I_X)) \cong \left(\frac{\sigma_Y^2}{m\bar{Y}^2} - \frac{\sigma_{YX}}{m\bar{Y}\ \ \bar{X}}\right)I_X \qquad (4.24)$$

These results are used for proving the following proposition

Proposition 2. Under the set of hypothesis of proposition 1 the approximate MSE of

$$\hat{R}_N = \frac{\bar{y}}{\bar{x}_N} = \frac{\bar{y}}{\bar{x}} + \frac{\bar{y}}{\bar{x}^2+\bar{x}}I_X = r + r^{*}I_X \qquad (4.25)$$

is

$$MSE(\hat{R}_N) \cong \left[\frac{\sigma_Y^2}{m\bar{Y}^2} + \frac{\sigma_X^2}{m\bar{X}^2} - 2\frac{\sigma_{YX}}{m\bar{Y}\ \ \bar{X}}\right] + \left[\left[\frac{2\sigma_Y^2}{m\bar{Y}^2} + \frac{1}{(\bar{X}^2+\bar{X})^2}\left[(\mu_{\bar{x}4} - \bar{X}^4) + (\mu_{\bar{x}3} - \bar{X}^3) + \frac{\sigma_X^2}{m}\left(1 - 2(\bar{X}^2 + \bar{X})\right)\right]\right.\right.$$
$$\left.\left. - 2\frac{1}{m}\left(\frac{1}{\bar{Y}}\right)\left(\frac{1}{\bar{X}^2+\bar{X}}\right)\left(\sigma_{\bar{y}\bar{x}^2} + \sigma_{yx} - \bar{Y}\sigma_x^2\right) - \frac{\sigma_{YX}}{m\bar{Y}\ \ \bar{X}}\right]I_X\right]$$
$$= M_C + M_N \qquad (4.26)$$

Note that, if $I_X = 0$ the classical sampling results on ratio estimators are obtained.

5. Numerical studies

Only one of the multiple theoretical challenges of developing neutrosophic counterparts of sampling models is considered in this paper. The estimation of a ratio when the auxiliary variable is neutrosophic poses a set of theoretical problems to be solved for other classes of ratio estimators.

With the aims of illustrating the behavior of proposal, data obtained from four real life problems are analyzed numerically. They are:

P1. 230 persons with ages in the interval 15-35 were questioned on the preferences for 5 songs. The reports (Y) were measured in a scale 1-10. The mean of the daily downloads of the songs in the last 30 days was the auxiliary variable X. The manager of a record company is the decision maker.

DOI: 10.5281/zenodo.4030309

P2. The farmers tax-report was the variable of interest Y and X was the last month tax-payment. The population size was 450. The consensus of the specialists of the state office performed the role of the decision maker.

P3. The fuel consumption report of a fleet of 76 vehicles in two consecutive months was measured. Y was the consumption in the actual month and X the report in the previous month. The owner of the enterprise acted as decision maker.

P4. The monthly incomes of 117 employees of an enterprise was Y and X was the total amount of operations with their credit cards. The owner of the enterprise acted as decision maker.

The researchers had a complete knowledge of Y and X. Hence was possible to compute the values of the involved parameters. $\hat{R}_N$ was obtained from the sample results and compared with the true value of R_N . B random samples of size m were selected from each population and the accuracy of the estimates was measured computing

$$\alpha_{Pj} = \frac{1}{B}\sum_{s=1}^{B}|r - R|_s + \frac{1}{B}\sum_{s=1}^{B}|r^* - R^*|_s I_X, = \alpha_{Cj} + \alpha_{Nj}\ j = 1,..,4 \quad (5.1)$$

Using the population information was computed (4.26) for each population

$$MSE(\hat{R}_N) \cong M_{Cj} + M_{Nj} = MSE_j; j = 1, \dots, 4 \quad (5.2)$$

See the results of the study in Table 1.

Table 1. Results of the Monte Carlo experiments

Population	m	B	α_{Cj}	α_{Nj}	α_{Pj}	M_{Cj}	M_{Nj}	MSE_j
1	25	1000	0,324	0,072-0,250	0,396-0,774	1,561	6,331-7,117	7.892-8,678
2	50	1500	1,920	0,400-0.652	2,320-2,552	5,452	0,851-1,807	6,303-7,259
3	15	2600	2,690	0,841-1,741	3,531-4,431	8,785	6,263-11,549	15,048-20,334
4	25	2000	1,130	0,757-0,965	1,887-2,095	1,873	6,022- 7.625	7,895-8,498

A lecture of the lines of the previous table suggests that the samples averaged an absolute difference between the estimate and the true value, which take values in the corresponding fifth column. The MSE of the methods appears in the eighth column. The decision makers fixed the corresponding I_X. They considered that was obtained a good description of the impreciseness of the estimates rules by their appreciations.

Acknowledgements The authors thank Professors F. Smarandache and Said Broumi for their encouragement for developing some ideas on sampling based on neutrosophic theory and to the referees, whose suggestions allowed improving seriously the quality of this paper.

Funding: "This research received no external funding"

Conflicts of Interest: "The authors declare no conflict of interest."

REFERENCES

[1] Ajay, D., J. Aldring, S. Abirami, D. Jeni Seles Martina, "A SVTrN-number approach of multi-objective optimisation on the basis of simple ratio analysis based on MCDM method"International Journal of Neutrosophic Science Vol. 5, pp. 16-28, 2020

[2] Alhabib , R., M. M. Ranna , H. Farah , A.A. Salama "Some Neutrosophic Probability Distributions", Neutrosophic Sets and Systems, Vol. 22, pp 30-38, 2018.

[3] Atanassov K. T. "Intuitionistic Fuzzy Sets, Physica–Verlag" , Heidelberg, New York, 1999.

[4] Aslam, M. “A new sampling plan using neutrosophic process loss consideration”. Symmetry, Vol.10, 132, 2018.

[5] Aslam, M. “Neutrosophic analysis of variance application to university students”. Complex & Intelligent Systems Vol. 5, pp 403–407 (2019a)

[6] Aslam, M. “Control Chart for Variance Using Repetitive Sampling Under Neutrosophic Statistical Interval System”. IEEE Access, (2019b)

[7] Aslam, M. “Design of the Bartlett and Hartley tests for homogeneity of variances under indeterminacy environment”. Journal of Taibah University for Science, Vol. 14 , pp 6–10, 2020.

[8] Aslam A., O. H. Arif, 1 and R. A. K. Sherwani “New Diagnosis Test under the Neutrosophic Statistics An Application to Diabetic Patients”. BioMed Research International , 2020a, https//doi.org/10.1155/2020/2086185.

[9] Aslam, A., O. H. Arif, and R. A. K. Sherwani “A new attribute sampling plan using neutrosophic statistical interval method”. BioMed Research International” Vol 2020, pp 1-6, 2020b,

[10] Bouza, C. N. and A. Alomari Ranked Set Sampling. 65 Years Improving The Accuracy In Data Gathering. Elsevier, ISBN 978-0-12-815044-3, 2019.

[11] Cacuango Almeida, G.; J. Cadena Morillo ; M. Carrillo Palacio “Discernimiento e inferencia de la reinserción social en Ecuador, basada en conjuntos de números de 2-tuplas”. Inv. Operacional, Vol. 41, pp 637-646, 2020.

[12] Crespo Berti L. A. “Application of the neutrosophic system to tax havens with a criminal approach”, International Journal of Neutrosophic Science , Vol. 5, pp. 91-106, 2020.

[13] Hanafy, I.M. , A.A.Salama and K. Mahfouz “Correlation of Neutrosophic Data. International Refereed Journal of Engineering and Science”, Vol. 1, pp 39-43, 2012.

[14] Hatip A. “The Special Neutrosophic Functions” International Journal of Neutrosophic Science , Vol. 4, pp. 104-116, 2020.

[15] Patro S.K. and F. Smarandache “The Neutrosophic Statistical Distribution, More Problems, More Solutions “. Neutrosophic Sets and Systems, Vol. 12, pp 73-79, 2016.

[16] Saqlain M. , A. Hamza and S. Farooq “Linear and Non-Linear Octagonal Neutrosophic Numbers: Its Representation, $\boldsymbol{\alpha - Cut}$ and Application”, International Journal of Neutrosophic Science ,Vol. 3, pp. 29-43, 2020

[17] Singh, S. “Advanced Sampling Theory with Applications. How Michael selected Amy”. Kluwer, Dordrecht 2003.

[18] Smarandache, F. Neutrosophy “A New Branch of Philosophy”, in Multiple valued logic, An International Journal, Vol. 8, pp 297-384, 2002.

[19] Smarandache F. “Neutrosophic set a generalization of the intuitionistic fuzzy sets” , In Proceedings of the third conference of the European Society for fuzzy logic and Technolgye, EUSFLAT, Septamper Zittau Geamany; Univ. of Applied Sciences at Zittau Goerlit Vol. 2, pp 141-146, 2003.

[20] Smarandache, F. “Introduction to Neutrosophic Measure, Neutrosophic Integral, and Neutrosophic Probability”, Sitech-Education Publisher, Craiova – Columbus, 2013.

[21] Smarandache, F. Introduction to Neutrosophic Statistics. Infinite Study. 2014.

[22] Smarandache, F. “Neutrosophic Overset, Neutrosophic Underset, and Neutrosophic Offset Similarly for Neutrosophic Over-/Under-/Off- Logic, Probability, and Statistics”, Pons Editions Brussels, Belgium, 2016.

[23] Subzar, M., C., N. Bouza, S. Maqbool, T A Raja and B. A. Para Robust ratio type estimators in simple random sampling using Huber m estimation Inv. Operacional Vol. 40, pp 201-209, 2019.

[24] Schweizer P. “The Natural Bases of Neutrosophy”, International Journal of Neutrosophic Science , Vol. 9, pp. 100-109, 2020

DOI: 10.5281/zenodo.4030309