

A NEW METHOD FOR DETERMINATION OF ADEQUATE SAMPLE SIZE

Mohammad Mousaei Sanjerehei¹ and Philip W. Rundel²

¹Department of Plant Protection, Yazd Branch, Islamic Azad University, Yazd,
Iran

²Department of Ecology and Evolutionary Biology, University of California
Los Angeles, California, United States of America
E Mail: mmusaei@iauyazd.ac.ir; rundel@biology.ucla.edu

Received October 19, 2018

Modified March 10, 2019

Accepted April 15, 2019

Abstract

Most population variables such as density seldom appear to be normally distributed. Therefore, the equations for normal distributions may overestimate the sample size required to obtain an accurate estimate of the variable. In this research, the total required sample size for obtaining the density with 5 and 10% precision was determined in random, uniform and clumped distribution patterns. A significant power relationship ($p < 0.0001$) was found between the actual sample size and the square of coefficient of variation for achieving 5 and 10% precision. The sample size calculated for normal distribution based on 95% confidence level and 5 and 10% precision was respectively 1 to 10 and 1 to 4 times the actual sample size obtained in the different distribution patterns. The overestimation increased with an increase in the values of coefficient of variation. A new method is presented based on population and sample coefficient of variation for estimating the required sample size. The new method can be used to obtain a reliable estimate of sample size using a wide range of data types in different study designs.

Key Words- Coefficient of Variation, Clumped, Density, Normal Distribution, Random, Sample Size, Uniform.

1. Introduction

Density is one of the important and readily obtainable attributes that is frequently estimated to describe the biological characteristics of a species or community. Density is defined as the number of individuals in a given unit of area or the reciprocal of the mean area of space per individual. Measurements of density are useful for determining species composition, diversity, richness, intra and inter specific competition and association, distribution patterns of organisms, studying population size and growth rate, monitoring ecosystem conditions as well as evaluating the influence of environmental factors on species and ecosystems (Bonham, 1989; Dacier et al., 2011; Mousaei Sanjerehei, 2011; Mousaei Sanjerehei et al., 2011; Olen et al., 2016; Chiang et al., 2017; Cowling et al., 2018). Some population attributes such as cover and biomass can be indirectly obtained through density estimates (McIntyre, 1953; Strong, 1966; Mousaei Sanjerehei and Basiri, 2008). Plot methods are among the simplest to use and have been frequently used to estimate the density of organisms (Becker and Crockett, 1973; Anderson and Marcus, 1993; Larson et al., 2008; Mousaei Sanjerehei and Basiri, 2008; Alwin et al., 2010; Mahajan and Fatima, 2017). Plot sampling may efficiently integrate variations in distribution pattern and departures from

randomness and may, therefore, be accurate to estimate density. The highest efficiency of density estimate is generally indicated by the plot that provides the smallest variance for density measurement.

One of the most important factors influencing the power of the density estimate is adequate sample size (number of sampling units). Appropriate sample size estimation generally depends on a variety of factors including the purpose of study, population size, level of precision and confidence, desired statistical power and degree of variability in the estimated attribute (Cochran, 1977; Mousaei Sanjerehei and Bassiri, 2007; Charan and Biswas, 2013; Eng, 2003; Malterud et al., 2016; Chow et al., 2017). Adequate sample size is usually obtained based on the sample mean, variance, desired confidence level and precision (Cochran 1977; Chow et al., 2017).

Coefficient of variation (CV) is a standardized measure of dispersion of a probability distribution or frequency distribution. The CV is defined as the ratio of standard deviation to mean and is a dimensionless measure that quantifies the degree of variability relative to the mean. It is one of the important and widely used measures for determination of sample size. Cochran (1977) developed one of the most widely used methods of sample size estimation based on CV for normal populations. Van Belle and Martin (1993) developed methods for known variance normal and the noncentral t distribution for estimating sample size based on coefficient of variation. Kelly (2007) presented an approach for determining the necessary sample size so that the estimated coefficient of variation accurately reflects the corresponding population value by achieving a sufficiently narrow confidence interval.

Variance of the number of individuals in sampling units (e.g., plots) is strongly affected by the pattern of dispersal of individuals. In the random dispersion, each sampling unit has an equal chance of hosting an individual and the resultant variance is close to mean ($\sigma^2/\mu = 1$). Where individuals have a clumped pattern, there are numerous sampling units with no individuals and a few with many individuals. In clumped patterns, the variance is much greater than the mean ($\sigma^2/\mu > 1$). In uniform distribution, most sampling units contain exactly the same number of individuals and as a result the variance is much less than the mean ($\sigma^2/\mu < 1$) (Ludwig et al., 1988; Mousaei Sanjerehei and Bassiri, 2007). Therefore, sample size needed for sampling of clumped communities is greater than that of random communities, which in turn is greater than that of uniform communities.

There are certain statistical frequency distributions that, because of their variance-to-mean properties, have been used as models of these types of distribution patterns: the Poisson distribution ($\sigma^2 = \mu$) for random patterns, the negative binomial ($\sigma^2 > \mu$) for clumped patterns and the (positive) binomial ($\sigma^2 < \mu$) for uniform patterns (Gardner et al., 1995; Lickfeldt et al., 2002; Castano-Meneses et al., 2003).

A variety of methods have been developed for estimating sample size for Poisson and binomial distributions. For example, Signorini (1991) presented sample size methods for Poisson regression based on various distributions of a single covariate, and a family of multivariate exponential-type distributions of multiple covariates. Zhu and Lakkis (2014) derived a sample size formula on the basis of likelihood function of

the negative binomial model with the ability to incorporate dispersion parameter and exposure time.

Commonly estimated population variables such as density are seldom normally distributed (Clark 2001; Lang 2004; Roraas et al., 2012). When data are skewed, the sample size estimated based on the variance and confidence intervals for normal populations is overestimated. Too big sample size can make the research unwieldy, wasting both time and effort. This necessitates introduction of the models which can reliably present an adequate sample size for accurately estimation of population attributes with different spatial patterns of data. The objective of this research was to develop the models for determining the required sample size for obtaining density with 5 and 10% precision in random, uniform and clumped distributions.

2. Materials and Methods

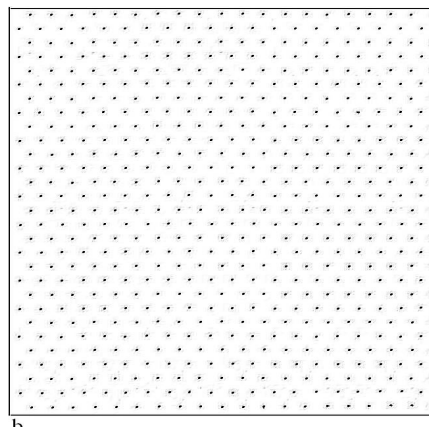
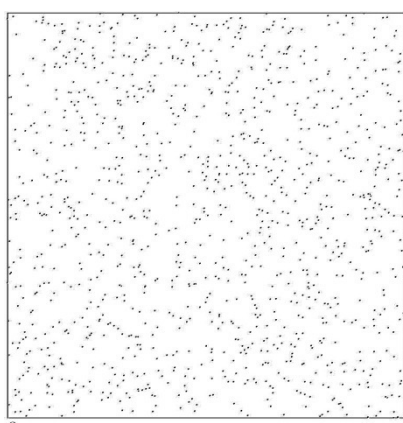
2.1. Sampling design

Total required sample size (number of sampling units) was determined for obtaining the density with ± 5 and $\pm 10\%$ precision in this research. Precision is the percentage difference between the estimated and true population density. Mean, standard deviation and density were determined by using square and rectangular different sized plots in the following distribution patterns of points;

A) Random distribution of points with the densities of 0.05, 0.1, 0.2, 0.5, 0.75, 1, 2, 4, 10.5, 19 and 35 points cm^{-2} (Fig. 1-a)¹.

B) Uniform (regular) distribution of points with the densities of 1, 2, 10 and 20 points cm^{-2} (Fig. 1-b).

C) Clumped distribution of points with different intensity of clumping and the densities of 0.1, 0.26, 1, 2, 5 and 9 points cm^{-2} (Fig. 1-c and d).



¹ - The whole sampling area of 580 cm^2 ($20 \times 29 \text{ cm}$) may be not fully covered by some plot sizes. For example $1 \times 1 \text{ cm}$ plots can cover the whole area, but $2 \times 2 \text{ cm}$ plots can cover an area of $20 \times 28 \text{ cm}$. Therefore the actual density calculated by the plot sizes that could not fully cover the area are close to the values presented above.

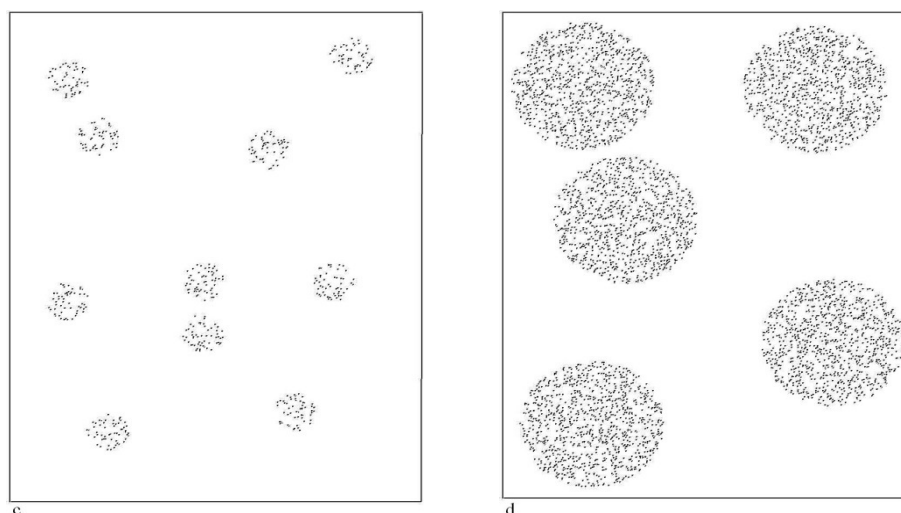


Figure. 1- a) Random distribution of points with a density of 2 cm^{-2} (a total 1160 points/580 cm^2), b) Uniform distribution of points with a density of 1 cm^{-2} , c and d) Clumped distribution of points with density of 1 and 5 points cm^{-2} and different intensities of clumping (ten 58-point and five 580-point clumps)

Random distribution of points with the density of 0.05 to 35 was generated on an A4-size paper ($20 \times 29 = 580 \text{ cm}^2$) using ArcGIS10.2. Each paper was divided into $0.5 \text{ cm} \times 0.5 \text{ cm}$ (0.25 cm^2) plots. Number of points in the 0.25 cm^2 plots were counted continuously column by column (Fig. 2) and the data were entered in Excel. The number of points in contiguous plots was summed to obtain the number of points in square and rectangular shaped plots with different sizes (1×1 , 1.5×1.5 , 2×2 , 2.5×2.5 , 3×3 , 4×4 , 5×5 , 0.5×1 , 1×3 , 1×5 , $1 \times 7 \text{ cm}$ and $2 \times 14 \text{ cm}$) (Fig. 2).

Since sampling of a 580 cm^2 area may be not enough for calculating the adequate sample size for achieving 5 and 10% precision (Lyon, 1968; Otypkova and Chytry, 2006), the data of plots for each size were replicated 20 times to estimate the density in an area of $580 \times 20 = 11600 \text{ cm}^2$. To obtain the density by each plot size, the data of each plot size were randomly reordered and the density was calculated by consecutively adding plot data (1^{st} , 1^{st} and 2^{st} , 1^{st} , 2^{st} and 3^{st} plot, *etc.*) as $\frac{\sum X_i}{NA}$, where X_i is the number of points in the plots, A is the area of plot and N is the number of plots (Table 1.). Mean and standard deviation of the number of individuals were also consecutively calculated by the plots.

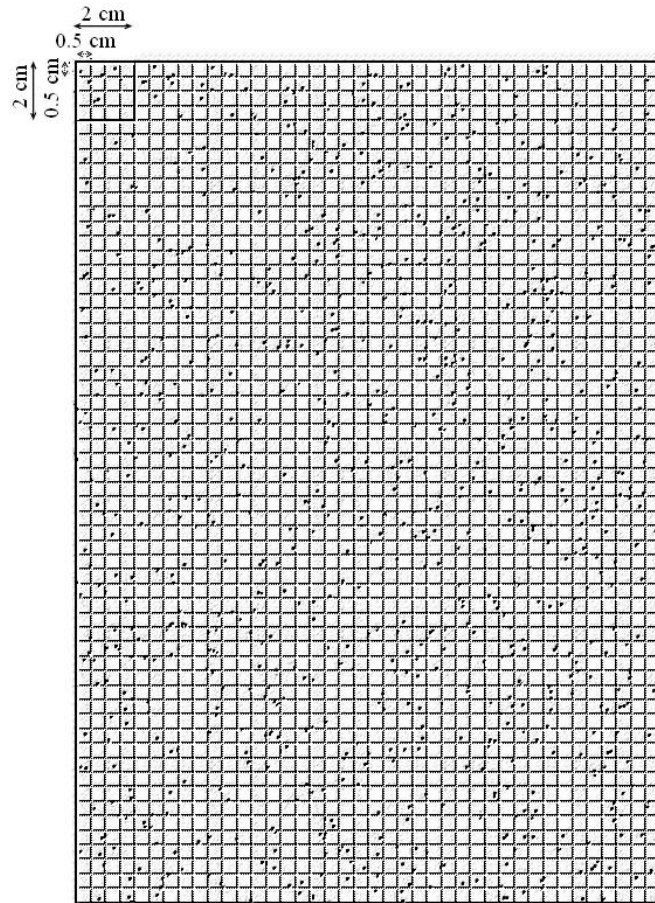


Figure. 2: Division of each paper into $0.5\text{cm} \times 0.5\text{cm}$ (0.25 cm^2) plots and counting the number of points in each plot. The figure shows combination of 16 contiguous 0.25 cm^2 plots for obtaining number of points in $2 \times 2\text{cm}$ (4 cm^2) plots.

Uniform distribution of points was generated using Point software on an A4-size paper. The density was then estimated using different sized plots as explained for random distribution.

Clumped distributions of points with different intensity of clumping (different number of clumps and number of points within clumps) and the density of 0.1 to 9 cm^{-2} were generated using ArcGIS. The Intensities of clumping generated were: ten 6-point, fifteen 10-point, ten 58-point, twenty 58-point, five 580-point and nine 580-point clumps.

The data for each plot size were replicated 20 times and randomly reordered to obtain the density as explained for random patterns.

2.2. Determination of the distribution of points and data

The spatial pattern of points was measured by Morisita's Index of Dispersion (ID_m) as $ID_m = N \frac{\sum x_i^2 - \sum x_i}{(\sum x_i)^2 - \sum x_i}$, where x is the number of individuals in the plots and N is number of plots. ID_m is equal to 1 for random distribution, <1 for uniform distribution and >1 where there is clumping (Morisita, 1962). Significant departures of ID_m from 1 (randomness) was tested using Chi-square test as $\chi^2 = ID_m(\sum x_i - 1) + N - \sum x_i$.

2.3. Determination of adequate sample size and development of models

The required sample size for obtaining the density with 5% (N_5) and 10% (N_{10}) precision were determined for each plot size and distribution pattern. N_5 was the sample size (number of plots) by which a precision of 5% was obtained and the precision remained below 5% as the number of plots increased (Table 1). N_{10} was the sample size by which a precision of 10% was obtained and the precision remained below 10% as sampling continued with more plots (Table 1).

For each precision, the type of relationship between the actual sample size and the population coefficient of variation ($\frac{\sigma}{\mu}$) was determined. σ (population standard deviation) is the standard deviation of the number of individuals in all plots for each size, and μ (population mean) is the mean number of individuals in the plots for each size = $\left(\frac{\text{total number of individuals in the plots}}{\text{total number of plots for each size}} \right)$. To test the models, RMSE (root mean square error) was calculated as $RMSE = \sqrt{\frac{(d_i - D_i)^2}{N-2}}$, where d_i is the predicted N_5 and N_{10} by the models and D_i is the observed N_5 and N_{10} .

To obtain the required sample size using sample standard deviation and mean, the relationship between the actual sample size and $\left(\frac{ts}{k\bar{x}} \right)$ was determined for $k=0.05$ and 0.10. s and \bar{x} are respectively the sample standard deviation and mean. t is determined from t -table based on the degree of freedom ($n-1$) at 95% and 99% confidence level.

AD	x	$\sum x$	A	N	ED	P	\bar{x}	s	$\frac{s}{\bar{x}}$	$t_{0.05}$	$t_{0.01}$	V_{5-95}	V_{5-99}	V_{10-95}	V_{10-99}
4	10	10	2.25	1	4.44	11.11	10	-	-	-	-	-	-	-	-
4	14	24	2.25	2	5.33	33.33	12	2.83	0.236	12.706	63.657	59.9	300.2	30	150.1
4	11	35	2.25	3	5.19	29.63	11.67	2.08	0.178	4.303	9.925	15.3	35.4	7.7	17.7
4	9	44	2.25	4	4.89	22.22	11	2.16	0.196	3.182	5.841	12.5	22.9	6.2	11.5
4	9	53	2.25	5	4.71	17.78	10.6	2.07	0.195	2.776	4.604	10.8	18	5.4	9
4	8	61	2.25	6	4.52	12.96	10.17	2.14	0.21	2.571	4.032	10.8	17	5.4	8.5
4	10	71	2.25	7	4.51	12.7	10.14	1.95	0.192	2.447	3.707	9.4	14.3	4.7	7.1
4	9	80	2.25	8	4.44	11.11	10	1.85	0.185	2.365	3.499	8.8	12.9	4.4	6.5
4	7	87	2.25	9	4.3	7.41	9.67	2	0.207	2.306	3.355	9.5	13.9	4.8	6.9
4	11	98	2.25	10	4.36	8.89	9.8	1.93	0.197	2.262	3.25	8.9	12.8	4.5	6.4
4	9	107	2.25	11	4.32	8.08	9.73	1.85	0.19	2.228	3.169	8.5	12.1	4.2	6
4	12	119	2.25	12	4.41	10.19	9.92	1.88	0.19	2.201	3.106	8.3	11.8	4.2	5.9
4	9	128	2.25	13	4.38	9.4	9.85	1.82	0.185	2.179	3.055	8.1	11.3	4	5.6

4	8	136	2.25	14	4.32	7.94	9.71	1.82	0.187	2.16	3.012	8.1	11.3	4	5.6
4	9	145	2.25	15	4.3	7.41	9.67	1.76	0.182	2.145	2.997	7.8	10.9	3.9	5.5
4	9	154	2.25	16	4.28	6.94	9.63	1.71	0.178	2.131	2.947	7.6	10.5	3.8	5.2
4	8	162	2.25	17	4.24	5.88	9.53	1.7	0.178	2.12	2.921	7.6	10.4	3.8	5.2
4	11	173	2.25	18	4.27	6.79	9.61	1.69	0.176	2.11	2.898	7.4	10.2	3.7	5.1
4	9	182	2.25	19	4.26	6.43	9.58	1.64	0.171	2.101	2.878	7.2	9.9	3.6	4.9
4	11	193	2.25	20	4.29	7.22	9.65	1.63	0.169	2.093	2.861	7.1	9.7	3.5	4.8
4	7	200	2.25	21	4.23	5.82	9.52	1.69	0.177	2.086	2.845	7.4	10.1	3.7	5
4	9	209	2.25	22	4.22	5.56	9.5	1.65	0.174	2.08	2.831	7.2	9.8	3.6	4.9
4	8	217	2.25	23	4.19	4.83	9.44	1.65	0.175	2.074	2.819	7.3	9.9	3.6	4.9
4	9	226	2.25	24	4.19	4.63	9.42	1.61	0.171	2.069	2.807	7.1	9.6	3.5	4.8
4	8	234	2.25	25	4.16	4	9.36	1.6	0.171	2.064	2.797	7.1	9.6	3.5	4.8
4	10	244	2.25	26	4.17	4.27	9.39	1.58	0.168	2.06	2.787	6.9	9.4	3.5	4.7
4	11	255	2.25	27	4.2	4.94	9.44	1.58	0.167	2.056	2.779	6.9	9.3	3.4	4.6
4	5	260	2.25	28	4.13	3.18	9.29	1.76	0.19	2.052	2.771	7.8	10.5	3.9	5.3
4	8	268	2.25	29	4.11	2.68	9.24	1.75	0.189	2.048	2.763	7.8	10.5	3.9	5.2
4	8	276	2.25	30	4.09	2.22	9.2	1.73	0.188	2.045	2.756	7.7	10.4	3.8	5.2
4	8	284	2.25	31	4.07	1.79	9.16	1.71	0.187	2.042	2.75	7.6	10.3	3.8	5.1
4	12	296	2.25	32	4.11	2.78	9.25	1.76	0.19	2.04	2.745	7.8	10.4	3.9	5.2
4	15	311	2.25	33	4.19	4.71	9.42	2	0.212	2.038	2.741	8.7	11.6	4.3	5.8

Table 1: Determination of the required sample size for estimating the density of 4 cm² with a precision of 5% by 2.25 cm² plots in the random distribution of points. N₅ was obtained to be 23. This is the sample size by which a precision of 5% was obtained and the precision remained below 5% as sampling continued with more plots. AD: actual density, x: number of individuals in plot, Σx: cumulative x, A: plot area, N: sample Size (number of plots), ED: estimated density, P: % difference between the actual and estimated density (%precision), \bar{x} : mean number of individuals in plots, s: standard deviation of the number of individuals in plots, t_{0.05}: t- table at a=0.05, t_{0.01}: t table at a=0.01. t-values start from the second row because of the degree of freedom n-1. $V_{5-95} = \frac{t_{0.05}s}{0.05\bar{x}}$, $V_{5-99} = \frac{t_{0.01}s}{0.05\bar{x}}$, $V_{10-95} = \frac{t_{0.05}s}{0.1\bar{x}}$, $V_{10-99} = \frac{t_{0.01}s}{0.1\bar{x}}$.

3. Results

3.1. Distribution of the points and data

Morisita's index of dispersion (ID_m) was 0.9-1.04 ($p > 0.05$) for the random patterns indicating a random distribution of points and an agreement with a Poisson series of data. ID_m for the uniform pattern scenarios was 0.2-0.85 ($p < 0.05$) suggesting the uniformity of dispersion and binomial distribution of the data. For the clumped patterns, ID_m was in a range of 1.41-2.78 ($p < 0.05$) indicating different intensity of clumping and fit of negative binomial model to the data.

3.2. Relationship between the actual sample size and the population coefficient of variation

By using different dispersion patterns and plot sizes and shapes, a wide range of variance (0.155-31686), mean (0.012-980) and coefficient of variation (0.037-6.25) was obtained. There was a significant power relationship ($p < 0.0001$) between the population coefficient of variation and the required sample size for obtaining the density with 0.05 and 0.1 precision (Figs. 3 and 4). RMSE of the sample size models for achieving 0.05 and 0.1 was 24.5 and 17 respectively. Based on the size of data for

dependent variable (sample size of 3-6242 for 0.05 precision model and sample size of 1-3872 for 0.10 precision model), the obtained RMSE can indicate the efficiency of prediction of the models.

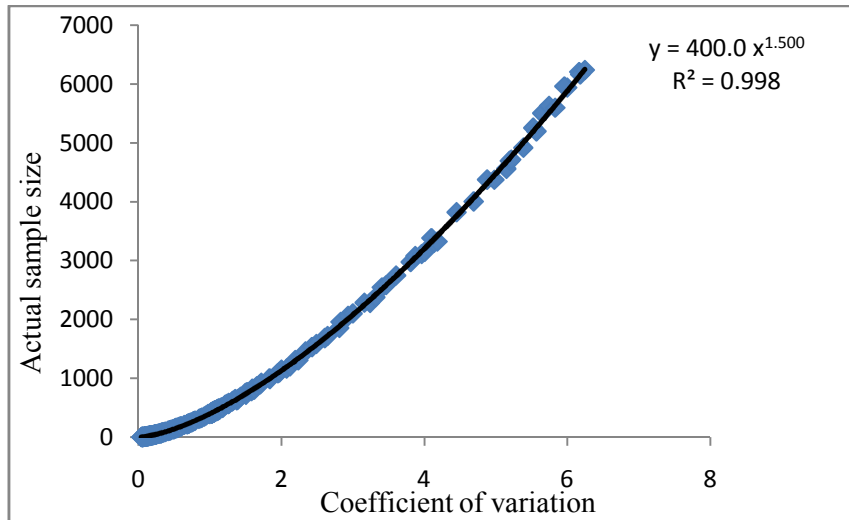


Figure. 3: The relationship between the actual sample size and the population coefficient of variation ($\frac{\sigma}{\mu}$) for achieving the precision of 0.05

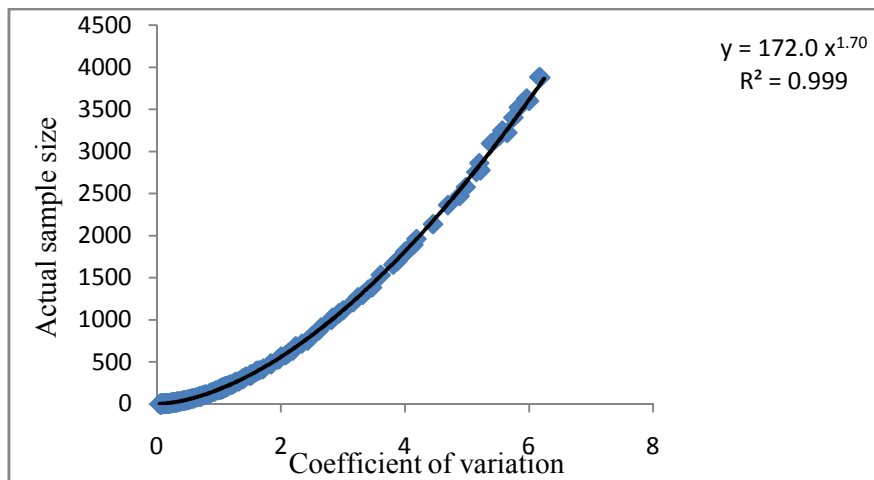


Figure. 4: The relationship between the actual sample size and the population coefficient of variation ($\frac{\sigma}{\mu}$) for achieving the precision of 0.10

3.3. Comparison between the actual sample size and the sample size for normal distribution

The sample size calculated using Cochran's equation for normal distribution $\left(\frac{ts}{k\bar{x}}\right)^2$ based on 0.05 precision and 95% confidence level was 1 to 10 times the actual sample size obtained using the studied designs (Fig. 5). For 0.10 precision, the sample size calculated by the equation was 1 to 4 times the actual sample size (Fig. 5). As shown in Figure 5, the overestimation of sample size by $\left(\frac{ts}{k\bar{x}}\right)^2$ equation was significantly correlated to the population CV and increased with an increase in the values of CV. There was also a significant correlation ($\alpha < 0.01$) between the actual sample size and $\left(\frac{ts}{k\bar{x}}\right)$ for $k = 0.05$ and 0.1 at 95 and 99% confidence levels.

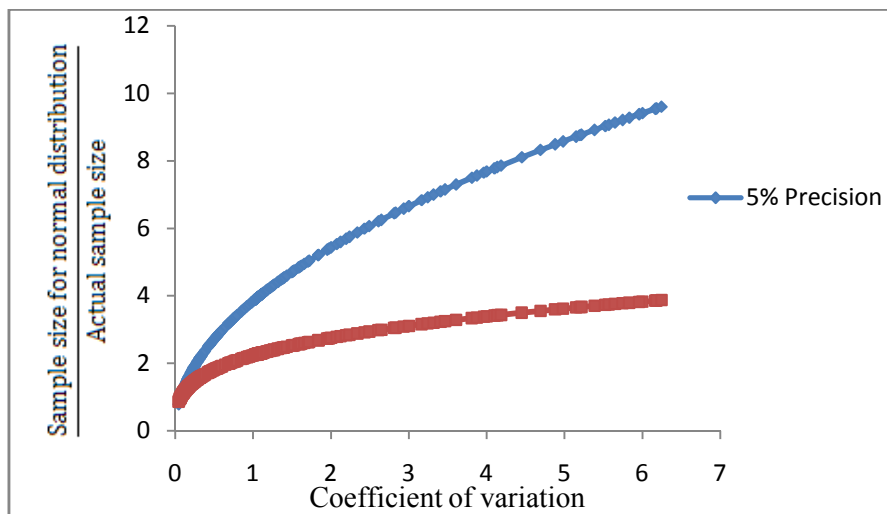


Figure. 5: The ratio of sample size calculated by Cochran's equation $\left(\frac{ts}{k\bar{x}}\right)^2$ for normal population to the actual sample size and its relationship to the population coefficient of variation $\left(\frac{\sigma}{\mu}\right)$ for achieving 5 and 10% precision

The models developed based on population and sample coefficient of variation for estimating an adequate sample size for the precision of 0.05 and 0.1 are presented in table 2. 95% upper and lower confidence intervals are presented for the mean of sample size.

The developed methods		
	Precision=0.05	Precision=0.10
	$N_5 = 400 \times \left(\frac{\sigma}{\mu}\right)^{1.5}$	$N_{10} = 172 \times \left(\frac{\sigma}{\mu}\right)^{1.7}$
95% Confidence intervals	$LCL_{95\%} = 386.4 \times \left(\frac{\sigma}{\mu}\right)^{1.5}$ $UCL_{95\%} = 413.9 \times \left(\frac{\sigma}{\mu}\right)^{1.5}$	$LCL_{95\%} = 166.1 \times \left(\frac{\sigma}{\mu}\right)^{1.7}$ $UCL_{95\%} = 177.9 \times \left(\frac{\sigma}{\mu}\right)^{1.7}$
Confidence level	Precision=0.05	Precision=0.10
95%	$N_5 = 1.63 \times \left(\frac{ts}{0.05\bar{x}}\right)^{1.5}$	$N_{10} = 1.093 \times \left(\frac{ts}{0.10\bar{x}}\right)^{1.7}$
99%	$N_5 = 1.082 \times \left(\frac{ts}{0.05\bar{x}}\right)^{1.5}$	$N_{10} = 0.687 \times \left(\frac{ts}{0.10\bar{x}}\right)^{1.7}$

Table 2- The developed methods based on population coefficient of variation (σ/μ) and sample coefficient of variation (s/\bar{x}) for determining an adequate sample size with precision of 0.05 and 0.10. N_5 and N_{10} are respectively the adequate sample size for achieving the precision of 5 and 10%. σ and μ are the population standard deviation and mean. s and \bar{x} are the sample standard deviation and mean. t is determined from t -table based on the degree of freedom ($n-1$) obtained from a preliminary sample and 95% and 99% confidence level. 0.05 and 0.10 are the precision for N_5 and N_{10} . LCL and UCL are respectively the upper and lower confidence level for the sample size mean.

4. Discussion

Many population variables such as density seldom follow a normal distribution. When data are skewed, the sample size equations that are based on standard deviation and confidence intervals for normal populations may overestimate the sample size required to accurately estimate the variable (Grerig 1964; Haukos et al., 1998; Piscard et al., 2004).

In this research, we determined the actual sample size for obtaining an accurate density using a wide range of data types based on population and sample coefficient of variation. The efficiency of CV for determination of sample size has been indicated in several studies (Cochran 1977; Van Belle and Martin 1993; Kelly 2007). A significant power relationship was found between the sample size and CV for achieving 5 and 10% precision (Ward et al., 2009; Taranets et al., 2012).

The sample size calculated by Cochran's method for normal population was greater than the actual sample size, and this overestimation increased with increasing values of CV. This is in line with the statement that when CV is very high (i.e., > 0.5)

the sample size estimation formulas for normal distributions lead to sample sizes that are generally greater than the exact ones and will increasingly overpower the study (Rasmussen et al 2019). We also found a significant correlation between the actual sample size and the sample CV indicating that sample CV can be an accurate estimator of population CV (Singh, 1986).

Although some data transformation methods such as logarithmic (Papanastasis, 1977; Curran-Everett, 2018), square root (Heady and Van Dyne, 1965; Malone, 1968; Hoda, 2018) and angular (Bartlett, 1947; Hinkley, 2014) transformations have been used for non-normal data in order to obtain symmetry, forcing the use of normal statistical methods on non-normal data to estimate adequate sample size does not yield the desired result, the valid use of data.

Other approaches such as the use of standard deviation of the consecutive means, instead of the standard deviation of the individual samples (Clark 2001), weighted sample size by dividing the variances into several groups with similar values (Kim et al., 2009) and control of the probability with which the desired margin of error is achieved (Gregoire and Affleck, 2018) have been recommended for calculation of sample adequacy based on skewed data. However use of these methods is efficient only when the form of the change of variance with mean level and the exact distribution of the original data are known (Cochran, 1977; Fitzner and Heckinger, 2010).

Kuno (1968) used the regression of mean crowding on mean density for estimating sample size, but stated that the model is valid only when sample mean is normally distributed about the true mean of the population. Cundill and Alexander (2015) presented a sample size equation based on calculations on the link function (log) scale for the negative binomial data, but they concluded that their method has little advantage for Poisson and binomial distributions.

The accuracy of sample size calculation obviously depends on the accuracy of the estimate of the parameters used in the calculations (Eng, 2003, 2007). Since the sample size models presented in this research were developed based on using the true population mean and standard deviation, they can be used to obtain a reliable estimate of sample size in a variety of research studies. In addition, the use of different types of data and a wide range of CV for making the models enable them to provide an efficient sample size in various distribution patterns.

References

1. Alwin, T.G., Fox, M.G. and Spence Cheruvellil, K.. (2010). Estimating lake-wide watermilfoil weevil (*Euhrychiopsis lecontei*) density: the roles of quadrat size, sample size, and effort. *Journal of Aquatic Plant Management (JAPM)*, 48, p.96.
2. Anderson, S. and Marcus, L.F. (1993). Effect of quadrat size on measurements of species density, *Journal of Biogeography*, p.421-428.
3. Bartlett, M.S. (1947). The use of transformations, *Biometrics*, 3(1), p.39-52.
4. Becker, D.A. and Crockett, J.J. (1973). Evaluation of sampling techniques on tall-grass prairie, *Journal of Range Management*, 26, p.61-65.

5. Bonham, C. D. (1989). *Measurements for Terrestrial Vegetation*, NY: John Wiley and Sons, 338.
6. Castano-Meneses, G., García-Franco, J.G. and Palacios-Vargas, J.G. (2003). Spatial distribution patterns of *Tillandsia violacea* (Bromeliaceae) and support trees in an altitudinal gradient from a temperate forest in Central Mexico, *Selbyana*, 24 (1) p. 71-77.
7. Charan, J. and Biswas, T. (2013). How to calculate sample size for different study designs in medical research?. *Indian Journal of Psychological Medicine*, 35(2), p.121-126.
8. Chiang, Y.C., Li, D. and Jane, H.A. (2017). Wild or tended nature? The effects of landscape location and vegetation density on physiological and psychological responses, *Landscape and Urban Planning*, 167, p.72-83.
9. Chow, Shein-Chung, Jun Shao, Hansheng Wang, and Yuliya Lokhnygina (2017). *Sample Size Calculations in Clinical Research*, Chapman and Hall/CRC Biostatistics Series.
10. Clark, D.L. (2001). Stabilization of the mean as a demonstration of sample adequacy, *Proceedings of America Society of Mining and Reclamation*, p.65-69.
11. Cochran, W.G. (1977). *Sampling Techniques*. John Wiley & Sons.
12. Cowling, R.M., Gallien, L., Richardson, D.M. and Ojeda, F. (2018). What predicts the richness of seeder and resprouter species in fire-prone Cape fynbos: Rainfall reliability or vegetation density?, *Austral Ecology*, 43, 614-622.
13. Cundill, B. and Alexander, N.D. (2015). Sample size calculations for skewed distributions, *BMC Medical Research Methodology*, 15(1), p.28 (9 pages).
14. Curran-Everett, D. (2018). Explorations in Statistics: the log transformation, *Advances in Physiology Education*, 42(2), p. 343-347.
15. Dacier, A., de Luna, A.G., Fernandez-Duque, E. and Di Fiore, A. (2011). Estimating population density of Amazonian titi monkeys (*Callicebus discolor*) via playback point counts, *Biotropica*, 43(2), p.135-140.
16. Eng, J. (2003). Sample size estimation: how many individuals should be studied?. *Radiology*, 227(2), p. 309-313.
17. Fitzner, K. and Heckinger, E. (2010). Sample size calculation and power analysis: a quick review, *The Diabetes Educator*, 36(5), p.701-707.
18. Gardner, W., Mulvey, E.P. and Shaw, E.C. (1995). Regression analyses of counts and rates: Poisson, overdispersed Poisson and negative binomial models, *Psychological Bulletin*, 118(3), p.392.
19. Gregoire, T.G. and Affleck, D.L. (2018). Estimating desired sample size for simple random sampling of a skewed population, *The American Statistician*, 72(2), p.184-190.
20. Haukos, D.A., Sun, H.Z., Wester, D.B. and Smith, L.M. (1998). Sample size, power and analytical considerations for vertical structure data from profile boards in wetland vegetation, *Wetlands*, 18(2), p.203-215.
21. Heady, H.F. and Van Dyne, G.M. (1965). Prediction of weight composition from point samples on clipped herbage, *Journal of Range Management*, 18, p. 144-148.
22. Hinkley, D.V. (2014). *Angular Transformation*, Wiley Stats Ref: Statistics Reference Online.
23. Hoda, M.N. (2018). The square root transformation, *International Journal of*

- Mathematical Archive EISSN 2229-5046, 9(4), p. 96-97.
24. Kelley, K. (2007). Sample size planning for the coefficient of variation from the accuracy in parameter estimation approach, *Behavior Research Methods*, 39(4), p. 755-766.
 25. Kim, K.Y., Chung, H.C. and Rha, S.Y. (2009). A weighted sample size for microarray datasets that considers the variability of variance and multiplicity, *Journal of Bioscience and Bioengineering*, 108(3), p.252-258.
 26. Kuno, E. (1968). Use of the regression of mean crowding on mean density for estimating sample size and the transformation of data for the analysis of variance, *Researches on Population Ecology*, 10(2), p.210-214.
 27. Lang, T. (2004). Twenty statistical errors even you can find in biomedical research articles, *Croatian Medical Journal*, 45, p. 361-370
 28. Larson, E.R., DiStefano, R.J., Magoulick, D.D. and Westhoff, J.T. (2008). Efficiency of a quadrat sampling technique for estimating riffle-dwelling crayfish density, *North American Journal of Fisheries Management*, 28(4), p. 1036-1043.
 29. Lickfeldt, D.W., Voigt, T.B. and Hamblin, A.M., 2002. Cultivar composition and spatial patterns in Kentucky bluegrass blends. *Crop science*, 42(3), pp.842-847.
 30. Ludwig, J.A., Quartet, L. and Reynolds, J.F. (1988). *Statistical Ecology: A Primer in Methods and Computing* (Vol. 1), John Wiley & Sons.
 31. Lyon, L.J. (1968). An evaluation of density sampling methods in a shrub community, *Journal of Range Management*, 21, p.16-20.
 32. Mahajan, M. and Fatima, S. (2017). Frequency, abundance, and density of plant species by list count quadrat method, *International Journal of Multidisciplinary Research*, 3(7), p.1-8.
 33. Malone, C.R. (1968). Determination of peak standing crop biomass of herbaceous shoots by the harvest method, *American Midland Naturalist*, 79, p.429-435.
 34. Malterud, K., Siersma, V.D. and Guassora, A.D. (2016). Sample size in qualitative interview studies: guided by information power, *Qualitative Health Research*, 26(13), p.1753-1760.
 35. McIntyre, G.A. (1953). Estimation of plant density using line transects, *Journal of Ecology*, 41(2), p. 319-330.
 36. Morisita, M. (1962). I σ -Index, a measure of dispersion of individuals, *Researches on Population Ecology*, 4(1), p.1-7.
 37. Mousaei Sanjerehei, M, and Basiri, M. (2008). Comparison and evaluation of density measurement methods on *Artemisia sieberi* shrublands in Yazd province, *Iranian Journal of Natural Resources*, 61, p. 235–251.
 38. Mousaei Sanjerehei, M. M. (2011). Determination of an appropriate quadrat size and shape for detecting association between plant species, *Ecological modeling*, 222(10), p. 1790-1792.
 39. Mousaei Sanjerehei, M.M. and Bassiri, M. (2007). comparison and evaluation of indices of dispersion patterns of plants on *artemisia sieberi* shrublands in Yazd province, *JWSS-Isfahan University of Technology*, 11(40), p.483-495.
 40. Mousaei Sanjerehei, M.M., Jafari, M., Mataji, A., Meybodi, N.B. and Bihamta, M.R. (2011). Facilitative and competitive interactions between plant species (an example from Nodushan rangelands, Iran). *Flora-Morphology*,

- Distribution, *Functional Ecology of Plants*, 206(7), p.631-637.
41. Olen, S.M., Bookhagen, B. and Strecker, M.R. (2016). Role of climate and vegetation density in modulating denudation rates in the Himalaya, *Earth and Planetary Science Letters*, 445, p.57-67.
 42. Otypkova, Z. and Chytry, M. (2006). Effects of plot size on the ordination of vegetation samples, *Journal of Vegetation Science*, 17(4), p.465-472.
 43. Papanastasis, V.P. (1977). Optimum size and shape of quadrat for sampling herbage weight in grasslands of northern Greece, *Journal of Range Management*, 30(6), p.446-448.
 44. Picard, N., Nouvellet, Y. and Sylla, M.L. (2004). Relationship between plot size and the variance of the density estimator in West African savannas, *Canadian Journal of Forest Research*, 34(10), p.2018-2026.
 45. Rasmussen, H. E., Ma, R., & Wang, J. J. (2019). Controlling type 1 error rate for sequential, bioequivalence studies with crossover designs, *Pharmaceutical Statistics*, 18(1), p. 96-105.
 46. Roraas, T., Petersen, P.H. and Sandberg, S. (2012). CIs and power calculations for within-person biological variation: effect of analytical imprecision, number of replicates, number of samples and number of individuals, *Clinical Chemistry*, 58(9), p. 1306-1313.
 47. Signorini, D. F. (1991). Sample size for Poisson regression, *Biometrika*, 78(2), p. 446-450.
 48. Singh, H. P. (1986). A note on the estimation of variance of sample mean using the knowledge of coefficient of variation in normal population, *Communications in Statistics-Theory and Methods*, 15(12), p. 3737-3746.
 49. Strong, C.W. (1966). An improved method of obtaining density from line-transect data, *Ecology*, 47(2), p. 311-313.
 50. Taranets, I. P., Smurov, A. V. and Kuznetsova, N. A. (2012). Different size sample representativeness in population accounting and estimation of horizontal distribution of soil-dwelling Collembola, *Moscow University Biological Sciences Bulletin*, 67(3-4), p. 126-129.
 51. Van Belle, G. and Martin, D. C. (1993). Sample size as a function of coefficient of variation and ratio of means, *The American Statistician*, 47(3), p. 165-167.
 52. Ward, M., Güntert, A., Campbell, J. and Pike, I. (2009). Proteomics for brain disorders—the promise for biomarkers, *Annals of the New York Academy of Sciences*, 1180(1), p. 68-74.
 53. Zhu, H. and Lakkis, H. (2014). Sample size calculation for comparing two negative binomial rates, *Statistics in Medicine*, 33(3), p. 376-387.