ISSN 1067-4136, Russian Journal of Ecology, 2020, Vol. 51, No. 1, pp. 64–71. © Pleiades Publishing, Ltd., 2020.

Adequate Sample Size and Appropriate Sampling Method for Estimating Grassland Production through Simulation

Nafiseh Fakhar Izadi[a], Kamal Naseri[a, *], and Mansour Mesdaghi[a]

[a]*College of Natural Resources and Environment, Ferdowsi University of Mashhad, Mashhad, Iran*

**e–mail: klnaseri@um.ac.ir*

Received June 19, 2018; revised January 29, 2019; accepted June 11, 2019

Abstract—One of the important factors in proper rangeland management is determination of grazing capacity which is related to accurate and precise estimation of forage production. As the estimation of forage production is only possible through sampling, so that determination of adequate sample size and suitable sampling method are the main objective of our applied research. The sampling was conducted in five locations in northern grasslands of Iran. In each location, two microplots of 64 m^2 were established based on international protocols. By combining the data of five locations, a layout of 640 m^2 was simulated in R software. This layout is considered as our statistical population. Primary sample sizes of 20, 25, 30, 35, and 40 were selected and by using different allowable errors with the lowest coefficient of variation, adequate sample sizes of 60 and 120 were obtained. These plot sizes were used to compare the accuracy and precision of methods including random, systematic, random–systematic, two-stage, and stratified sampling. The results showed as the allowable errors are decreasing, the sample sizes are increasing. In general, the mean of stratified sampling is almost equal to the actual mean of the population, indicating its higher accuracy and precision in comparison to other methods. Sample size of 60 and stratified sampling method with 3 to 5 strata were our ultimate selections.

Keywords: grassland, range production, sampling method, accuracy, precision, sample size

DOI: 10.1134/S106741362001004X

INTRODUCTION

Rangelands are an important part of the natural environment and the most extensive terrestrial ecosystem. They are important in supplying goods (such as forage, medical plants, and honey, etc.) and services (such as soil conservation, recreation, air and water purification, etc.) for human societies [1–3]. Based on above advantages, it is important to determine the grazing capacity correctly in order to prevent rangeland degradation. Consequently, a precise and accurate way to estimate forage production is required.

In many cases, measuring the total production of the study area is impossible, so sampling must be employed. Sampling reduces the amount of field operations and study budget. For range specialist, vegetation sampling is a tool to make an estimation of population parameters from samples. Sampling should be accurate and precise and give an unbiased estimate of the actual parameters. One way to evaluate precision is the replication of sampling and comparing the values of obtained samples [4, 5].

Sample size is the number of sampling units for estimating characteristics of a population [6–8]. The proper sample size can be calculated by the following equation [9–12, 5]. The formula of $n = (t \times s)^2/\mathrm{E}^2$ (derived from the t formula: $t = (\bar{Y} - \mu)/(s/\sqrt{n})$), where t is table value in Student's t-test, s is standard deviation and can be obtained by a primary sampling. E is the half-width of the desired confidence interval. Another version of above equation has been peresented as $n = (t \times CV)^2/\mathrm{A}^2$; where CV is the coefficient of variation and $A = E/\bar{Y}$ is allowable error expressed as a percentage of the mean [9].

Furthermore, vegetation structure and its spatial distribution have a significant impact on the size and shape of plots [13]. For estimating vegetation parameters (e.g. canopy coverage, production, and density), selection of plot sizes is very important for increasing the efficiency of sampling, reducing costs, and saving time [14, 15].

Plot size of any shape can affect the precision and accuracy of collected data. In general, a large number of small or medium plot sizes reflects changes in vegetation properly, and are thus preferred to a small number of large plot sizes [14, 16].

It has long been recognized that sampling procedures play an important role in population and community studies in ecology [7, 17]. Selection of an appropriate method of sampling in vegetation ecology

Table 1. Characteristics of study locations in northern regions of Iran

Study location (Province)	Longitude, m	Latitude, m	Elevation, m	Rainfall, mm	Mean production, kg/ha
Zarchak (Gilan)	50.071716	46.916799	2000	600	649
Spandol(Gilan)	82.019409	2.433286	1970	625	963
Torogh (Khorasan Razavi)	87.175007	5.693972	1240	262	1268
Dasht (Northern Khorasan)	56.055972	37.317803	1090	290	1954
Aselme (Khorasan Razavi)	88.403988	4.879084	1720	322	3331

is of great importance, since the selected samples should be representative of the desired population [6, 7, 9, 17–19].

Sampling methods must be chosen in accordance with the particular type of vegetation parameters being studied, i.e. plant density, size, and height [20].

It is clear that a weak and inefficient design of the sampling method can lead to loss of time, budget, and incorrect management decisions. Selection of different methods of sampling depends on several factors including heterogeneity of the population, variation among individuals, and inferences about the population. Available time and budget can also play an important role in making the final choice [9, 14, 16, 21].

There are several methods to establish sample plots; including random, systematic, random–systematic, stratified, and two-stage sampling. Definitely, the choice of sampling method will affect the accuracy and precision of estimates. Accuracy is related to the closeness of estimates to the population parameters and precision describes the extent to which measurements are scattered around their mean value [22, 23]. In other words, the lesser, a smaller the standard deviation (or standard error) the higher, the precision.

To estimate desired parameters, some studies on different sampling methods and plot sizes have been conducted. The number of kangaroos were estimated by using the simulation of different sampling methods and plot sizes in a 12 × 12 km grid of 144 plots of 1 km^2 [24]. To estimate the number of kangaroos, different methods of sampling including random with and without replacement, limited random and systematic sampling with large and small plots, and transects perpendicular to the gradient of kangaroo abundances were used [24].

In a study estimating herbaceous biomass [25], the relationship between the precision of number of samples, sampling density, spatial sampling techniques based on the spatial heterogeneity of biomass, and sampling procedures were investigated. Virtual plant communities were simulated based on the gamma distribution. The researchers concluded that the number of samples are a more influential factor than sampling density with respect to the precision of herbaceous biomass estimation. In addition, they confirmed that biomass heterogeneity strongly affects precision.

The advantage of our research is that by having the actual population parameters, we can control the estimated statistics with actual ones. In addition, using the powerful R software facilitates the simulation of sampling from our population.

The objective of this study is to simulate a small statistical population with known parameters of mean and standard deviation. By simulation based on the normal distribution, our aim is to find the relationships among precision and accuracy of sampling methods with different sample sizes. Based on our simulation for sampling production in different grassland communities, we can present an applicable methodology.

MATERIALS AND METHODS

Study Area

To collect production data, five grasslands were selected in northern regions of Iran (Table 1).

Sampling

Based on an international protocol [26] two macroplots of 64 square meters (8 × 8 m) were established in each location (Fig. 1). Then, in each location, forage production of 64 one square meter plots were clipped, air dried, and weighed.

A macroplot of 640, one square meter was built from two replications of 64 m^2 (2 × 64 m^2 = 128) plots at 5 locations, and was considered as the statistical population for this study. As the adjacent plots inside of 8 × 8 m were autocorrelated, we employed random sampling to reshuffle the values of each 64 plots. These reshuffling were repeated for 5 locations.

To determine the required sample sizes, 5 random plots were selected from original population of 640 plots and the mean and standard deviation were calculated. Then, a second, a third, a fourth and so on of 5 plots were added to the samples already drawn and again the same statistics were calculated.

The relationship between sample sizes and calculated statistics would be sketched for one hundred replications. Then, the required sample sizes (n) for our finite population (N) were calculated by using the following formula [10]):

$$n = \frac{1}{\left(\frac{A}{t \times CV}\right)^2 + \frac{1}{N}}.$$

Where: CV is the coefficient of variation, $A = E/\bar{Y}$ is allowable error expressed as a percentage of the mean, and N was the number of actual finite population of 640 plots.

The assumptions of approximate normal distribution are required to calculate sample size at the level of $\alpha = 0.05$ [9, 27].

The accuracy and precision of different sampling methods including random, systematic, random–systematic, two-stage, and stratified sampling were evaluated based on calculated sample sizes with 100 replications.

Furthermore, based on a new strategy of stratification, the actual population is divided to one stratum = STR1 (all locations, 640 plots), two strata = STR2 (Torogh–Dasht and Aselme–Zarchak–Spandol, with 256 and 384 plots, respectively), three strata = STR3 (Torogh–Dasht, Aselme, and Zarchak– Spandol, with 256, 128, and 256 plots, respectively), four strata = STR4 (Torogh, Dasht, Aselme, and Zarchak–Spandol, with 128, 128, 128, and 256 plots, respectively), and five strata = STR5 (Torogh, Dasht, Aselme, Zarchak, and Spandol, and each stratum with original number of 128 plots).

The initial plots of sampling for STR1 to STR5 were 20, 25, 30, 35, and 40 which their distribution in STR1 through STR4 were based on proportion to the area of each stratum. In STR5, the distribution of initial plots to be allocated to each stratum is computed by expressing each product of "area×standard deviation" as a proportion of the product sum. The allocated initial plots in each stratum were sampled with 100 replications.

Statistical Analysis

The production data of five sampling methods with calculated sample sizes were to be compared by using ANOVA. Significant methods were compared by using Tukey's Studentized range (HSD) which compare sampling methods while controlling the simultaneous confidence level [28]. The same procedure of analysis was applied to STR1 through STR5.

All data processing including simulation and data analysis were performed in R software. The R packages of sp, splancs, plotrix , and plyr were used in laying out the sampling methods. The packages of gmodels, and car were used for descriptive statistics and in ANOVA.

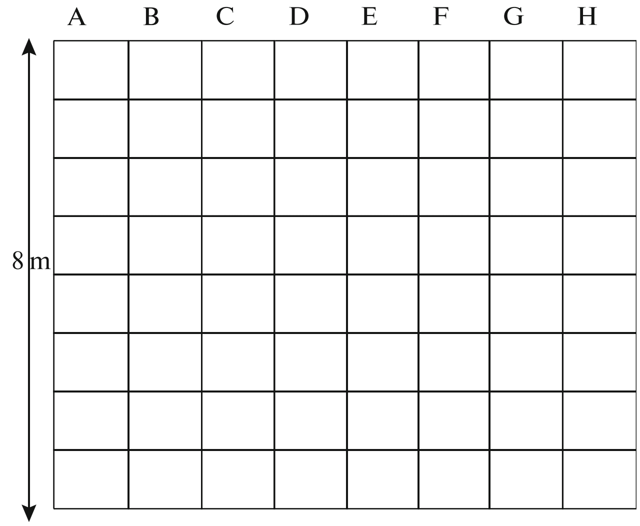

Fig. 1. Grids of 8 × 8 m plots.

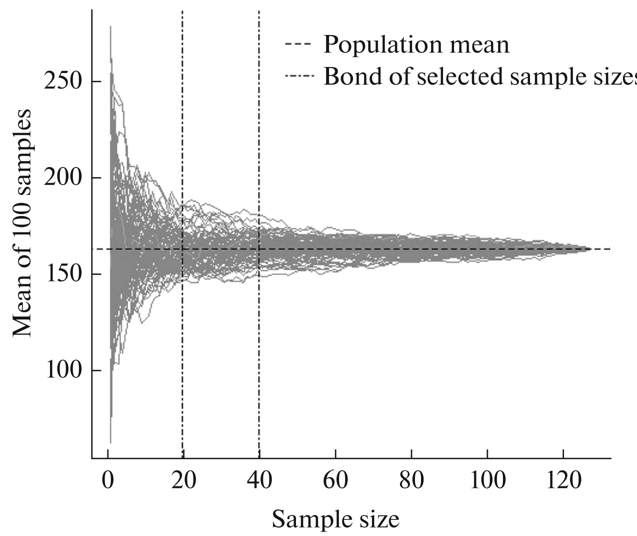

Fig. 2. As the size of 128 sample means of cumulative 5 plots drawn from the original population grows, the accuracy of sampling increases as the means moves closer to the true population mean. The required sample sizes were bonded between 20 and 40 scattered lines of mean.

RESULTS

Required Sample Size

The relationships among 128 sample means and sample sizes of 5, 10, and so on with 100 replications are shown in Figure 3. We have selected our initial number of plots that bonded between 20 and 40 sample sizes. As illustrated in Fig. 2, regardless of the skewed distribution of the original population, the scattered lines of all possible sample means were normally distributed [27, 28].

The result of required sample sizes of different initial plot sizes and different allowable errors was shown

Table 2. The required sample sizes based on 20 to 40 initial plots and 5 to 40 percent of allowable errors

A, %	Required sample sizes based on standard deviation of primary sampling				
	20 initial plots $CV = 63.08$	25 initial plots $CV = 62.85$	30 initial plots $CV = 62.15$	35 initial plots $CV = 59.16$	40 initial plots $CV = 59.42$
5	333	333	321	303	301
10	137	137	128	**117**	**117**
15	69	69	64	**58**	**58**
20	41	40	38	34	34
25	27	26	25	22	22
30	19	18	17	16	15
35	14	14	13	12	11
40	11	10	10	9	9

in Table 2. The *CV*'s were stabilized by increasing the initial plots numbers. By increasing A, the allowable errors, the number of required sample sizes were decreased.

Further analysis of comparing the precision and accuracy of different sampling methods were based on required samples size of 60 and 120 plots with 10 and 15 percent of allowable errors (the sample sizes of 117 and 58 were rounded to 60 and 120 to be divisible to actual population size) while the *CV*'s were stabilized (Table 2).

The Accuracy and Precision of Sampling Methods

Figure 4 illustrates 100 confidence intervals of five different sampling methods which were built on required sample sizes of 60 and 120. Intervals not containing the population mean are highlighted by + signs at their ends. In our simulation of 100 samples, less than 5 intervals out of 100 (5 percent) would be expected not to contain the actual population mean. The narrower the bond of confidence limits, the more precise was sampling method (Fig. 3).

The number of miss confidence intervals of stratified sampling is least and its bonds were narrowed to compare to the other methods to the actual mean. So it is more precise and accurate than the other methods (Fig. 3). In this regard, the random— systematic sampling method is the next choice. As shown in Fig. 3, with increasing the sample size from 60 to 120, the confidence intervals were narrowed.

By plotting different combination of five stratified methods against standard errors, the trends were sharply dropped in the first two strata and then were almost stabilized (Fig. 4).

We have run the ANOVA with 100 replications to compare different sampling methods and almost 80 and 60 percent of p—values were not significant ($p > 0.05$) for 60 and 120 sample sizes, respectively.

Simulation of Sample Selection in Different Sampling Methods

For random sampling 60 and 120 plots were selected from the actual population. For a systematic sampling of 60 and 120, odd numbers of plots of the actual population were chosen systematically. A typical plan of simulated random and systematic sampling for size of 60 is shown in Fig. 5. The common formulas of random and systematic sampling were employed for calculating statistics [5, 10].

For random—systematic sampling, 10 strip—plots from the 20 × 32 grid were systematically selected and then in each strip 6 and 12 one—m^2 plots were randomly chosen. The formulas for calculating statistics in this method are the same as two-stage sampling [10].

In two-stage sampling, first, 4 × 5 subplots were constructed and then 20 subplots were selected randomly. Then, 3 and 6 plots were selected randomly in each subplot. The formulas for this method were based on [22].

A typical plan of random—systematic and two-stage sampling methods for sample size of 60 is shown in Fig. 6.

For stratified sampling, first, the actual population is divided into 5 strata (the original number of plots in each location), and optimum allocation of number of plots (60 and 120) were based on standard deviation of each stratum. A typical simulation plan of stratified layouts with five strata and with two strata sampling methods for sample size of 60 plots is shown in Fig. 7.

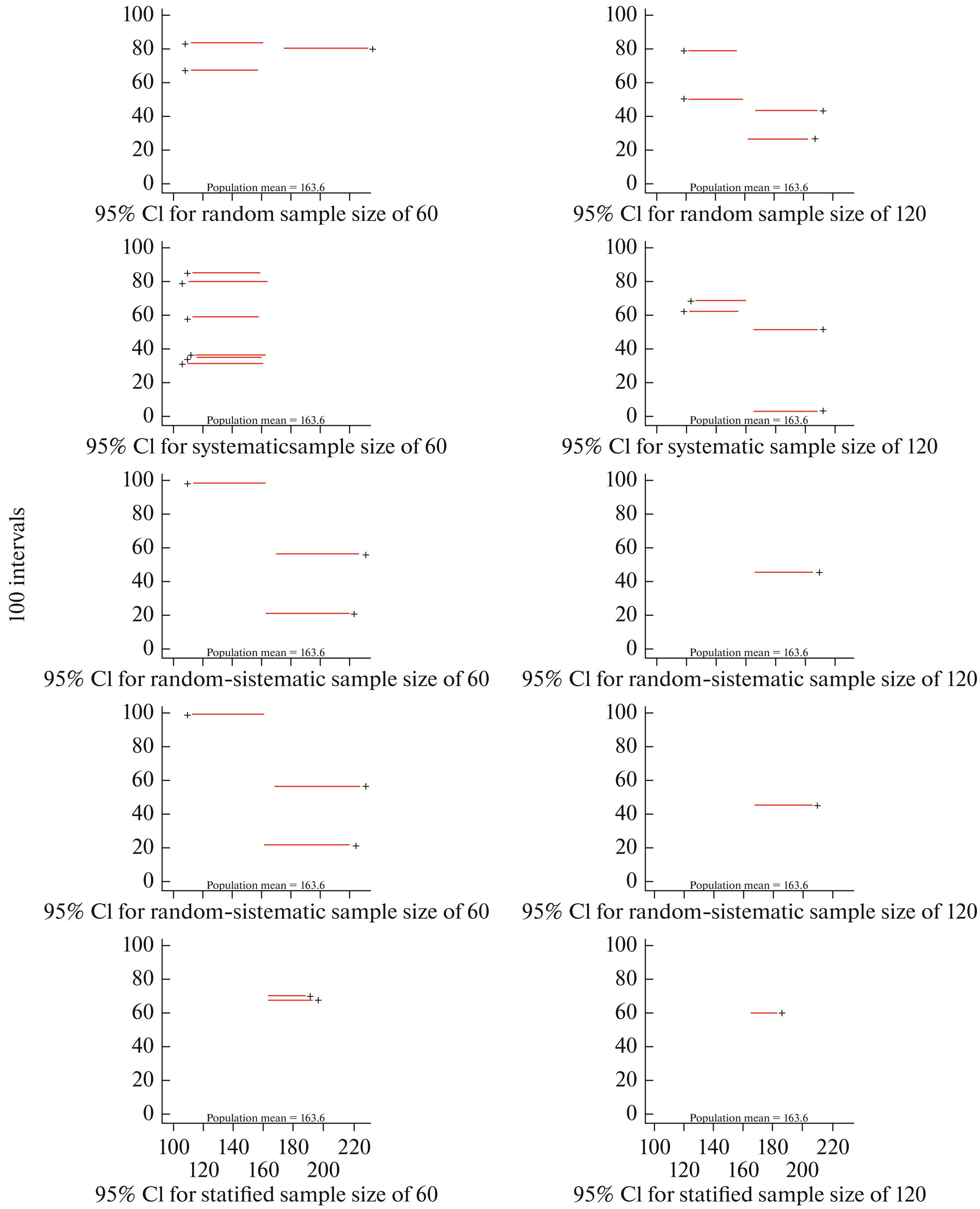

Fig. 3. One hundred 95 percent confidence intervals from random, systematic, random–systematic, two-stage, and stratified sample sizes of 60 and 120. Highlighted intervals miss the population mean.

DISCUSSION

Our strategy of simulated sampling in a population with known mean and variance was to apply the similar sampling procedures in an actual situation with a population of unknown parameters. In the real situation with high heterogeneity of vegetation, stratifying the study area would insure the first step of sampling [9, 17] As expected, the result of real stratified sampling method of five locations was accurate with the highest precision compared to other sampling methods (Fig. 4). So, selecting proper strata would be the

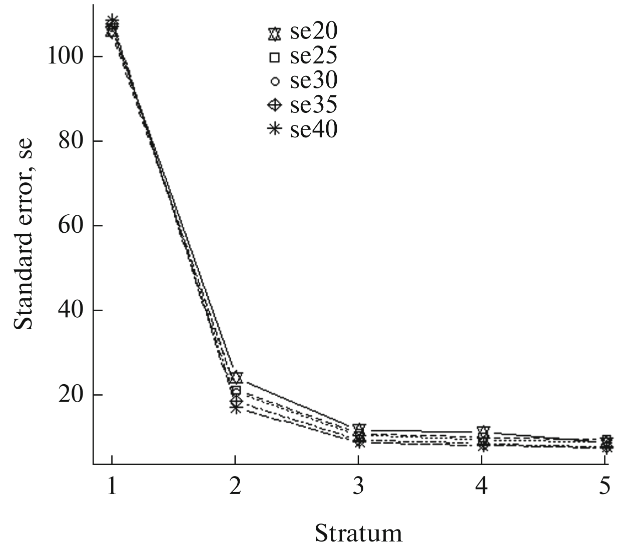

Fig. 4. One hundred replications for comparing standard errors of five stratified methods with different initial plots.

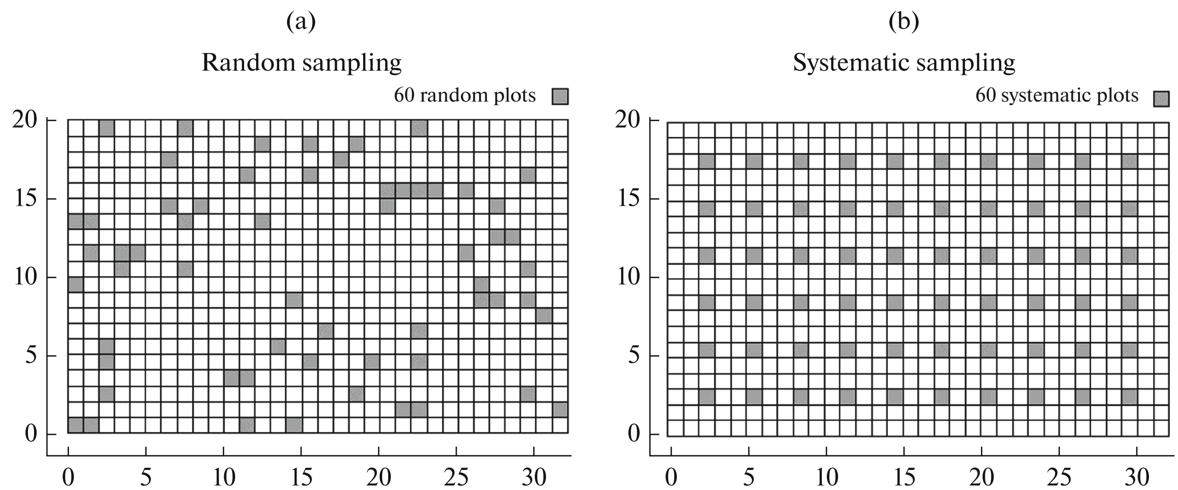

Fig. 5. A typical plan of simulation layouts of (a) random and (b) systematic sampling methods for sample size of 60.

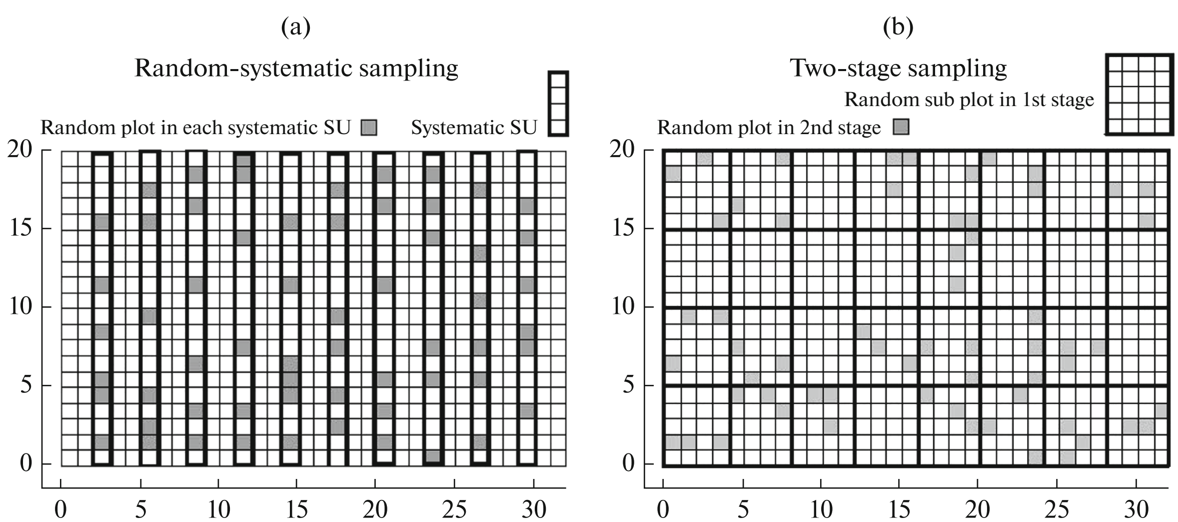

Fig. 6. The typical plan of simulation layouts of (a) random–systematic and (b) two-stage sampling methods for sample size of 60.

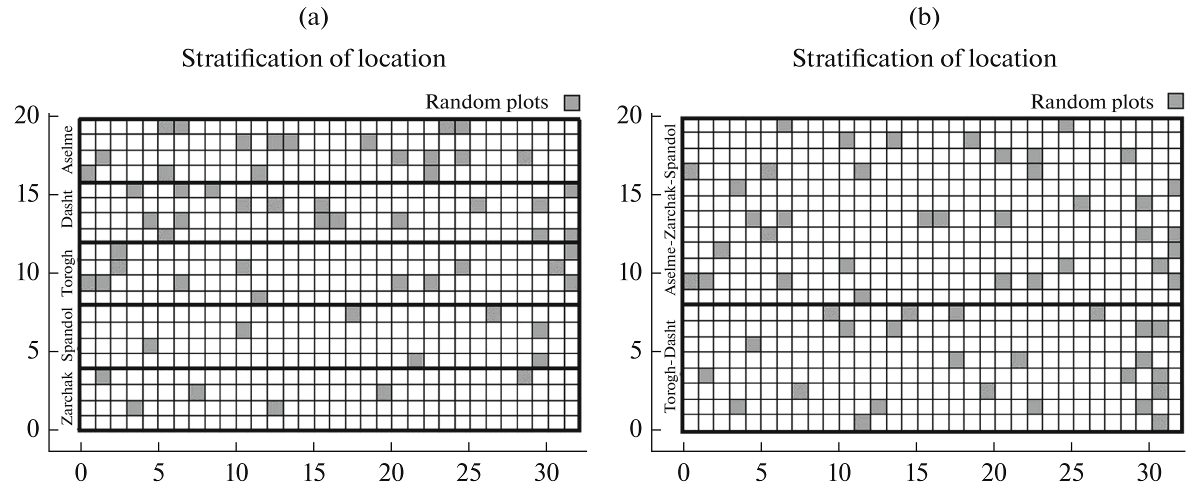

Fig. 7. The typical plan of simulation layouts of (a) stratified with five and (b) with two strata (STR2) sampling methods for sample size of 60.

first step of sampling the real populations [5, 22]. Although high accuracy and precision still belongs to the original five strata (Fig. 4), our result also indicated that stratification with different artificial strata did give high precision at three and four strata (Fig. 4).

In the next step, determination of sample and plot sizes are to be the most critical decision in sampling, but our plot size was already pre–selected of 1 m^2. Selection of proper sample size depends on the spatial variations of vegetation, budget, time, and expected precision; therefore, range specialists can sample adequate plots with proper sample sizes and allowable sampling error [14]. From standpoints of cost and time consumptions, finding minimum sample sizes which give us accurate and precise estimate of parameters is important. Selection sample size of 60 or 120 depends on allowable errors, sampling cost, and available budget [9, 10, 22] (For this reason, sampling costs were included in calculation of sample size allocations [5, 22].

On the other hand, non–significance of 80 and 60 percent of ANOVA's for our sample sizes of 60 and 120 respectively, reveals that even with lower sample sizes of 60, selection of the proper sampling procedure is only a matter of achieving precision. So, the sample size of 60 and stratified sampling method with 3 to 5 strata would be our ultimate selections. Selection of sample sizes less the 60 may be an alternative choice, but the production data of those sizes should be approximately normally distributed [28].

CONCLUSION

To determine adequate sample size and suitable sampling method, sampling was conducted in five locations in northern grasslands of Iran. By combining the data of five locations, a layout of 640 m^2 was simulated in R software. By using primary sample sizes and different allowable errors, adequate sample sizes were obtained. These plot sizes were used to compare the accuracy and precision of different methods. In general, the mean of stratified sampling is almost equal to the actual mean of the population, indicating its higher accuracy and precision in comparison to other methods.

ACKNOWLEDGMENTS

The Research of the Corresponding Author is Supported by a Grant from Ferdowsi University of Mashhad, Iran (N.3/32375).

COMPLIANCE WITH ETHICAL STANDARDS

We, as the author of this paper declare that we have no conflict of interest regarding to our paper. All of the collected plants are not endangered species and our clippings were not damaged the plants.

REFERENCES

1. Vermeire, L.T., Ganguli, A.C., and Gillen, R.L., A robust model for estimating standing crop across vegetation types, *J. Range Manag.,* 2002, vol. 55, pp. 494–497.
2. Flombaum, P. and Sala, O.E., A non–destructive and rapid method to estimate biomass and above ground net primary production, *J. Arid Environ.,* 2007, vol. 69, pp. 352–358.
3. Holechek, J.L., Pieper, R.D., and Herbel, C.H., *Range Management: Principle and Practices*, 6th ed., Upper Saddle River, NJ: Prentice Hall, 2011.
4. Bonham, C.D. *Measurement for Terrestrial Vegetation,* 2nd ed., New York: Wiley, 2013.
5. Cochran, W.G. *Sampling Techniques.* 3rd ed., New York: Wiley, 1977.

6. Thompson, S.K., *Sampling,* New York: Wiley, 2002.
7. Greig–Smith, P. *Quantitative Plant Ecology,* 3rd ed., London: Butterworth, 1983.
8. Gauch, H.G., *Multivariate Analysis of Community Data,* Cambridge: Cambridge Univ. Press, 1982.
9. Krebs, C.J., *Ecological Methodology,* 2nd ed., Menlo Park, CA: Benjamin Cummings, 1999.
10. Avery, T.E. and Burkhart, H. *Forest Measurement,* 4th ed., New York: McGraw–Hill, 2002.
11. Bartlett, J. E., Kotrlik, J.W., and Higgins, C.C., Organizational research: Determining appropriate sample size in survey research, *Inf. Technol. Learn. Perform. J.,* 2001, vol. 19, no. 1, pp. 23–31.
12. Elliott, J.M., *Some Methods for the Statistical Analysis of Samples of Benthic Invertebrates,* Freshwater Biological Association Sci. Publ. 25, 1983.
13. Cheryl, D., Raphael, P., and Claude, G., Comparative efficiency and accuracy of variable area transects versus square plots for sampling tree diversity and density, *Agrofor. Syst.,* 2010, vol. 79, pp. 223–236.
14. Cook, C.W., and Stubbendieck, J. *Range Research: Basic Problems and Techniques,* Lawrence, KS: Society for Range Management, 1986.
15. Walther, B.A., and Moore, J.L., The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance, *Ecography,* 2005, vol. 28. pp. 815–829.
16. Chambers, J.O., and Brown, R.W., *Method for Vegetation Sampling and Analysis on Revegetated Mined Lands,* Ogden, UT: USDA Forest Service, Intermountain Forest and Range Experiment Station, General Tech. Report INT–151, 1983.
17. Green, R.H., *Sampling Design and Statistical Methods for Environmental Biologists,* New York: Wiley, 1979.
18. Kent, M., *Vegetation Description and Data Analysis: A Practical Approach,* 2nd ed., New York: Wiley, 2012.
19. Van der Maarel, E., Vegetation ecology: An overview, in *Vegetation Ecology,* Van der Maarel, E. and Franklin, J., Eds., Oxford: Blackwell, 2005, pp. 1–3.
20. Barbour, M. G., Burk, J. H., Pitts, W.D., and Schwartz, F.S., *Terrestrial Plant Ecology,* 3rd ed., Addison Wesley Longman, 1999.
21. Huebner, C.D., Detection and monitoring of invasive exotic plants: A comparison of four sampling methods, *Northeast. Nat.,* 2007, vol. 14. pp. 183–206.
22. Freese, F., *Elementary Forest Sampling,* Agricultural Handbook no. 232, USDA Forest Service, 1962.
23. Elzinga, C.L., Salzer, D.W., and Willoughby, J.W., Measuring and Monitoring Plant Populations, *BLM Technical Reference* no. 1730–1, Denver, CO: Bureau of Land Management, 1998·
24. Fryxell, J.M., Sinclair, R.E., and Caughley, G., *Wildlife Ecology, Conservation, and Management,* 3rd ed., Boston, MA: Blackwell, 2014.
25. Tsutsumi, M., Itano, S., and Shiyomi, M., Number of samples required for estimating herbaceous biomass, *J. Rangeland Ecol. Manag.,* 2007, vol. 60, pp. 447–452.
26. Fraser, L.H., Jentsch, A., and Sternberg, M., What drives plant species diversiy? A global distributed test of unimodal relationship between herbaceous species richness and plant biomass, *J. Veget. Sci.* 2014, vol. 25, pp. 1160–1166.
27. Glantz, S.T., *Primer of Biostatistics,* 7th ed., New York: McGraw–Hill, 2012.
28. Steel, R.G.D., Torrie, J.H., and Dickey, D.A. *Principles and Procedures of Statistics: A Biometrical Approach,* 3rd ed., New York: McGraw–Hill, 1997.