Research Article

On the Efficiency of Some Alpha Diversity Indices: A Simulation Study Using Bootstrap Resampling

Jan Nino Gabayran Tinio[*]**, Carl Jericho Sebual**

Caraga State University, Department of Mathematics, Butuan City, Philippines
*Email: jgtinio@carsu.edu.ph

Abstract

It is essential to measure biodiversity to determine the stability of a community of species. A good measure gives precise, reliable, and efficient information about the community structure. This study was conducted to evaluate the statistical properties and test some alpha diversity indices' efficiency through a bootstrap resampling. The available birds' data were being applied in some alpha diversity indices. Moreover, the Shapiro-Wilk test is being used to test the normality of the index. All indices analyzed have p-values higher than the alpha level of significance (α=0.05), indicating that the index's bootstrap sampling distribution is normally distributed. Mostly, all indices are consistently efficient at large sample sizes. The Menhinick richness index is more efficient in measuring birds' species richness since it has the smallest means squared error (MSE) values. Among the diversity indices studied in this paper, Simpson's diversity index generated smaller mean squared error (MSE) values; thus, it is a more efficient diversity index in measuring birds' diversity. The Shannon evenness index is more efficient among the five evenness indices in measuring species evenness of birds because it provides the smaller MSE values.

Keywords: Alpha indices, biodiversity, biodiversity indices, bootstrap resampling, efficiency

Introduction

Efficiency is a measure of an estimator's quality, and it can be measured by empirically comparing variations of relative values of several possible estimators (Nikulin, 1994). A good estimator

provides a close estimation of the parameter that is being estimated and produces an estimate that has the quality of being efficient. Efficiency will serve as a basis to select an estimator that will closely capture the real scenario. An estimator that has a smaller variance would be preferred to be more efficient. The lower the estimator's variation, the more efficient the distribution of the estimator around the parameter being estimated (Trapani & Horvat, 2016). It is necessary to measure the efficiency of an estimator to determine the most accurate estimates for given statistical procedures and to evaluate the quality of an estimate when it is used to represent the whole sample or in making an inference about the unknown parameters (Fisher, 1921; Rao, 1962).

Efficiency can also be measured in biodiversity. In this study, diversity indices will serve as estimators, and these indices' efficiencies will be measured. Biological diversity, commonly known as biodiversity, is the scientific term for the diverseness of life on Earth. It also composes all living organisms from all sources, including wildlife, marine, and forests (Chivian & Bernstein, 2010). In ecology, a diversity index is a statistic that is used to quantify the species diversity in an ecosystem. These indices measure the number of different species present and the abundance of each species. They also give meaningful information about a species' commonness and rarity in a specific community (Gross et al., 2019). Biodiversity can be measured and monitored at three spatial scales, namely, alpha diversity, beta diversity, and gamma diversity. This study focuses on alpha diversity. Alpha diversity can be measured according to species richness, species evenness, or species diversity (Launchbaugh, 2009).

Species richness refers to the number of species present in a particular area or a community (Brown et al., 2007). Species evenness is the equitability in the distribution of individuals among species (Odum, 2000). Moreover, species diversity involves both richness and evenness of species (Krebs, 1999). The standard alpha diversity indices used nowadays are the Shannon index, Simpson index, and Berger-Parker index. Biologists have been extraordinarily innovative and proposed a large number of diversity indices. It is significant to understand the proper use of these measurements or indices by looking at its assumptions, different properties, and how they behave when it is applied on a particular community (Gross et al., 2019). It is essential to analyze and determine some alpha diversity indices' efficiency since an efficient index provides more reliable, precise, and informative results. This means that one can determine which diversity index is appropriate for a particular species or community structure.

To compare the efficiency of indices using bootstrap resampling, this study was conducted. Bootstrap resampling is a recently developed technique in making certain statistical inferences that require modern computer power to simplify intricate calculations of traditional statistical theory. Bradley Efron introduced it in 1979 and a very general sampling procedure for estimating the distributions of statistics based on independent observations. Moreover, bootstrap resampling is shown to be successful in many situations, which is being accepted as an alternative to the asymptotic methods (Efron & Tibshirani, 1993).

The study's main objective was to measure the efficiency of some alpha diversity indices using bootstrap inference on bird data. Specifically, it aimed to compute actual values of the different indices using avifauna or birds data and conduct and analyze simulation study to measure some alpha diversity indices' efficiency using bootstrap resampling.

Material and methods

Bootstrapping for Biodiversity Indices

In measuring the efficiency of the identified alpha diversity indices, the bootstrap simulation was conducted to the bird's data, as shown in figure 1. The data used in this study was extracted from the paper of Donna Mariel T. Calimpong and Olga M. Nuñeza entitled "Avifaunal Diversity of Bega Watershed, Prosperidad, Agusan del Sur." This set of data contains 83 species of birds, and it has 582 individuals. Five hundred eighty-two (582) individuals of birds were bootstrapped following the sample sizes, and several resamples used in this study. This study's sample sizes are 10, 30, 65, and 95; each having resamples of 50, 500, and 1000. Each sample and resample is used to solve the values of the different alpha diversity indices, and the corresponding mean, standard deviation, and mean squared error was calculated.

Also, it is necessary to test the normality of the sampling distribution of the bootstrap estimates. Shapiro-Wilk test is being applied to test the normality of the bootstrap estimates. This is necessary since the bootstrap estimates can only be meaningful when its sampling distribution is normal.

To evaluate the bootstrap estimates' efficiency, the mean squared error (MSE) is being considered. Mean squared error is used to compare the effectiveness of the estimators. If the mean squared error value is low, both its bias and variance are low; then an estimator is said to be efficient.
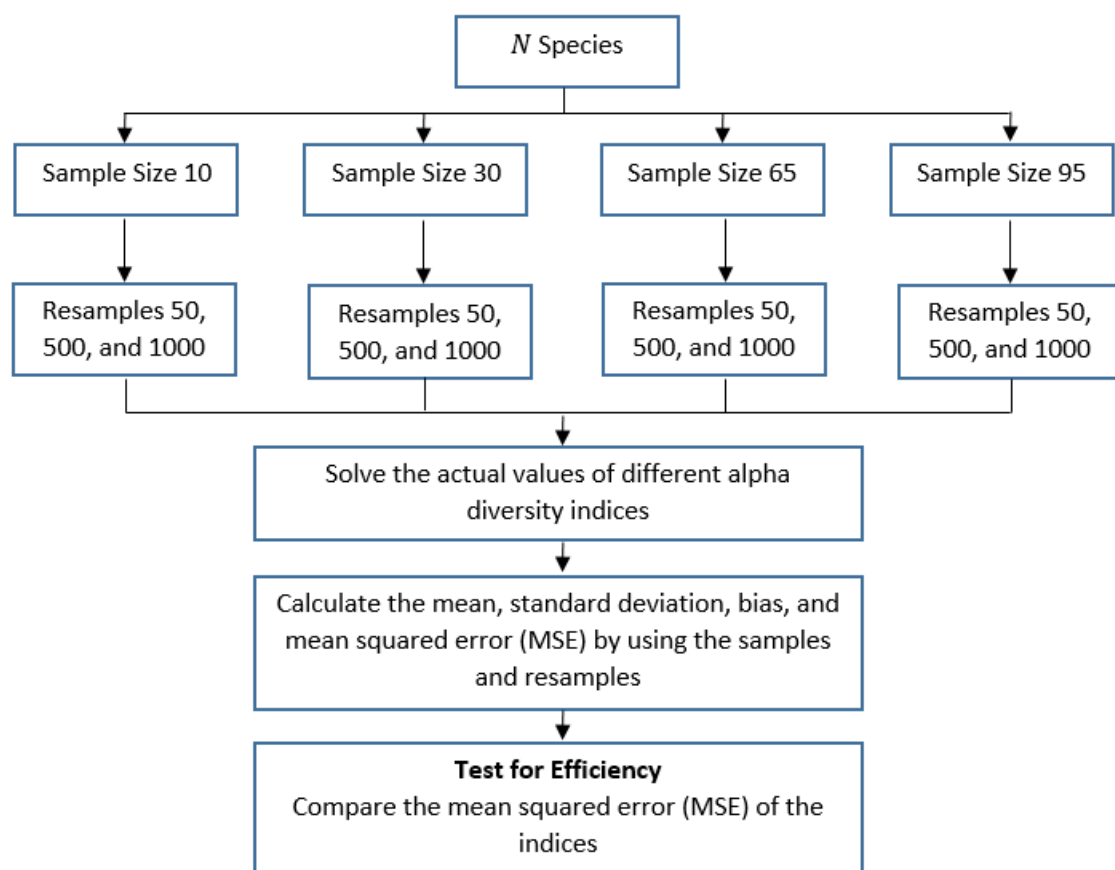

Figure 1. Bootstrap procedure for biodiversity Indices Flowchart

Results

Species Richness Measures

Figure 2 shows the bootstrap estimates of the species richness indices of Margalef and Menhinick.

Using the Shapiro-Wilk test, both Margalef and Menhinick richness indices are normally distributed since they produced p-values higher than the level of significance α=0.05 (See Table A.1).

In comparing the MSE values of both indices, the Menhinick richness index generated smaller MSE values than Margalef, considering every combination of sample and resample. This result also indicates that Margalef richness index produces larger variation when the sample size is small, making it sensitive to smaller sample sizes. Thus, the Menhinick richness index is more efficient in measuring the species richness of birds.
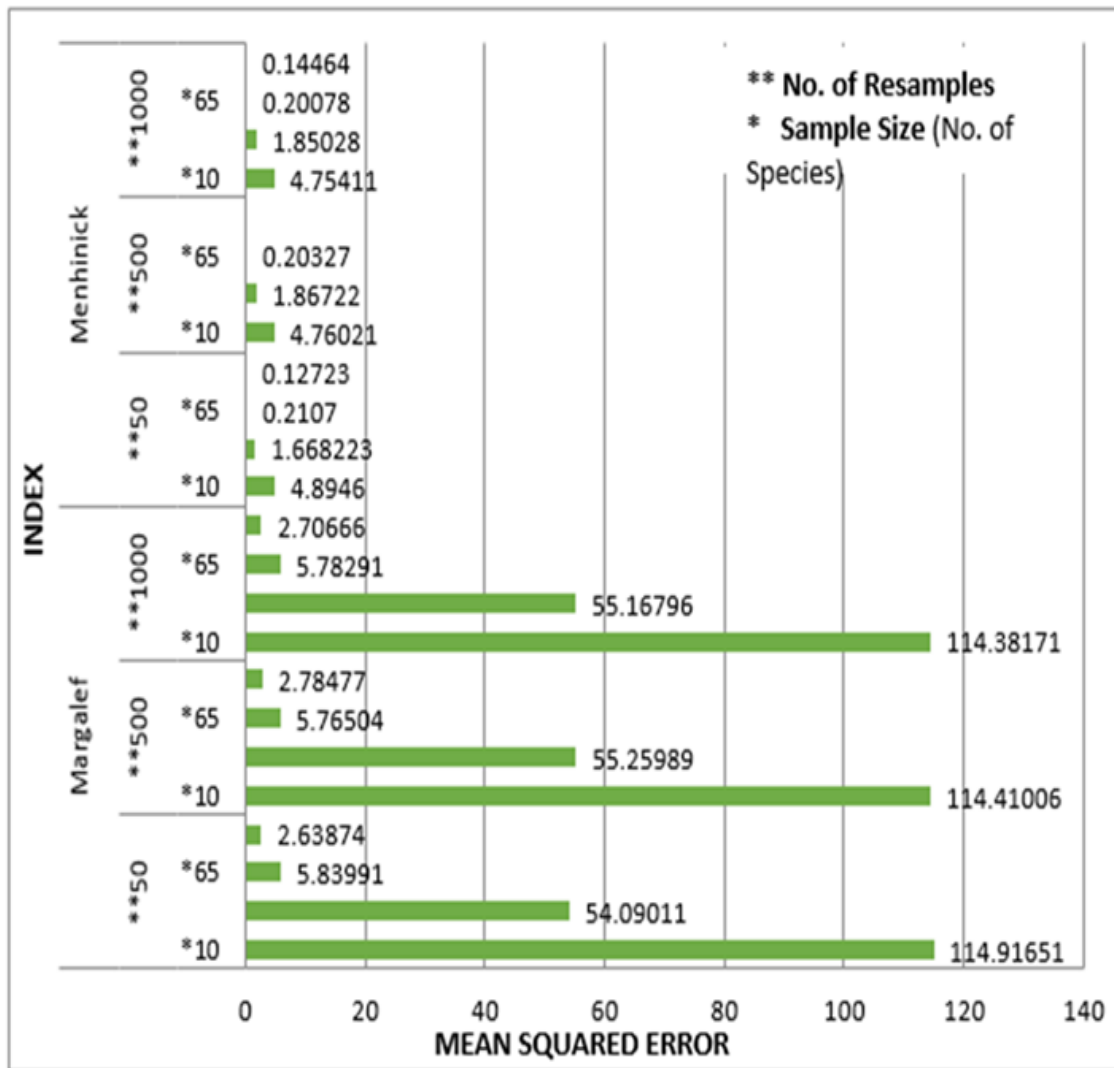

Figure 2. Bootstrap estimates for species richness indices

Species Diversity Indices

Figure 3 shows the bootstrap estimates of the diversity indices of Shannon, Simpson, and Berger-Parker. Using the Shapiro-Wilk test, Shannon, Simpson, and Berger-Parker diversity indices are normally distributed since they produced p-values higher than the level of significance α=0.05 (See Table A.1).

By comparing the MSE among the three indices, the Simpson diversity index produces smaller MSE values. This result suggests that the Simpson diversity index provides more practical and appropriate information for measuring the diversity of species of birds. Furthermore, it can also be

noted from this result that the Shannon diversity index is sensitive to smaller sample sizes since it produces higher variations in smaller sample sizes.
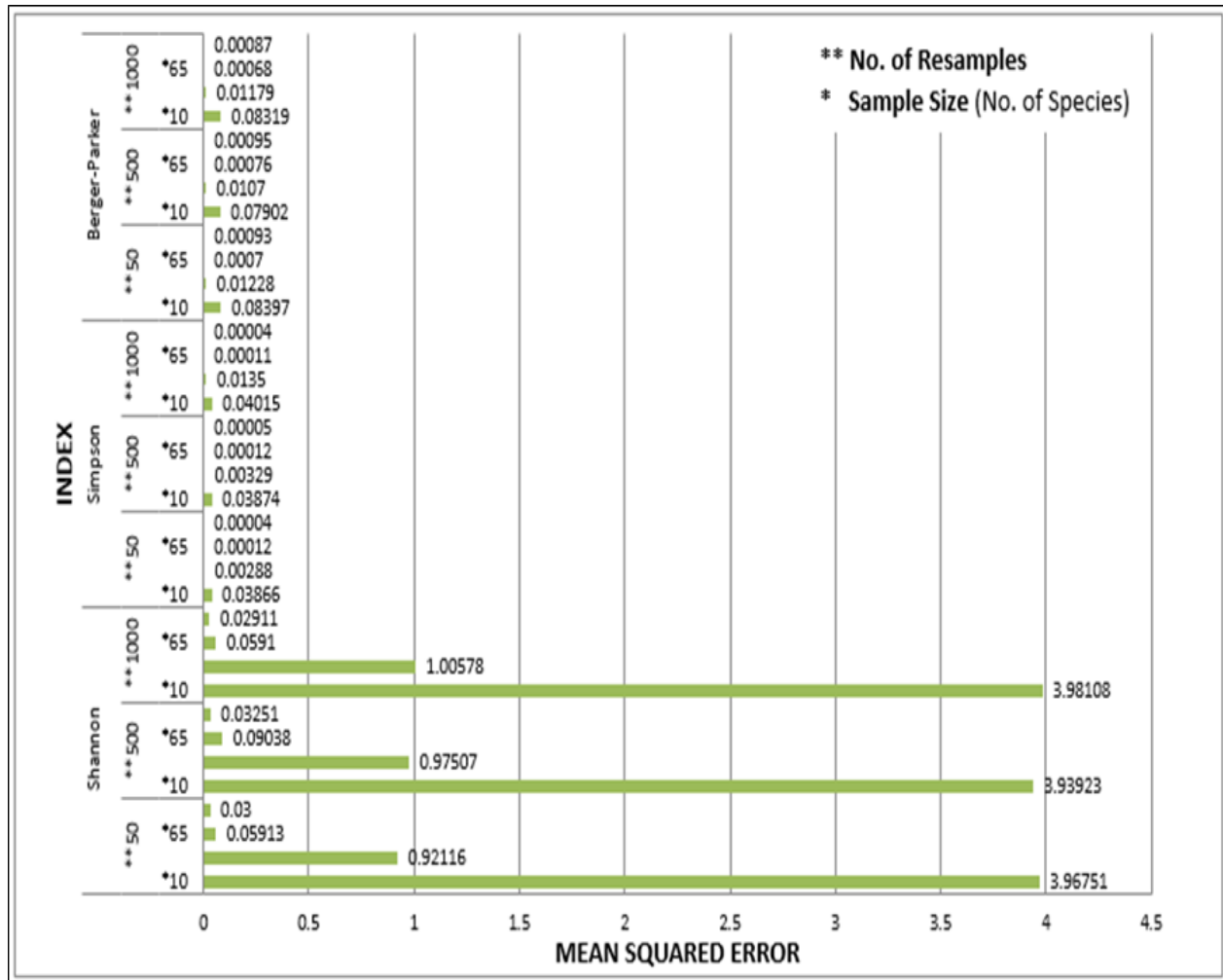

Figure 3. Bootstrap estimates for species diversity indices

Species Evenness Indices

Figure 4 shows the bootstrap estimates of the species evenness indices, namely, Shannon, Heip, Simpson, Hill's Ratio, and Modified Hill's Ratio. Using the Shapiro-Wilk test, the bootstrap estimates of these species evenness indices are normally distributed since each generated p-values greater than the level of significance α=0.05 (See Table A.1).

In comparing the MSE of the species evenness indices, Shannon's evenness index generated smaller MSE values. This suggests that the Shannon evenness index is more efficient compared to the other identified species evenness indices. This result also indicates that Simpson, Hill's Ratio, and Modified Hill's Ration evenness indices are sensitive to sample size. They produce higher variations when the sample sizes are smaller.

Discussion

In measuring the species richness of birds, the result of this paper suggests that the Menhinick richness index is more efficient than the Margalef richness index. The Menhinick index will estimate the richness of birds' species more accurately compared to the Margalef richness index. According

to Antoney (2015), the Menhinick richness index assumes a relationship between the number of species and the number of individuals, and Menhinick values vary with samples containing different numbers of individuals. He stated that the Menhinick index would be useful in measuring the species richness of the avian community. Besides, according to Williams et al. (2005), the Menhinick index is easy to compute, but it is sensitive to sample size.
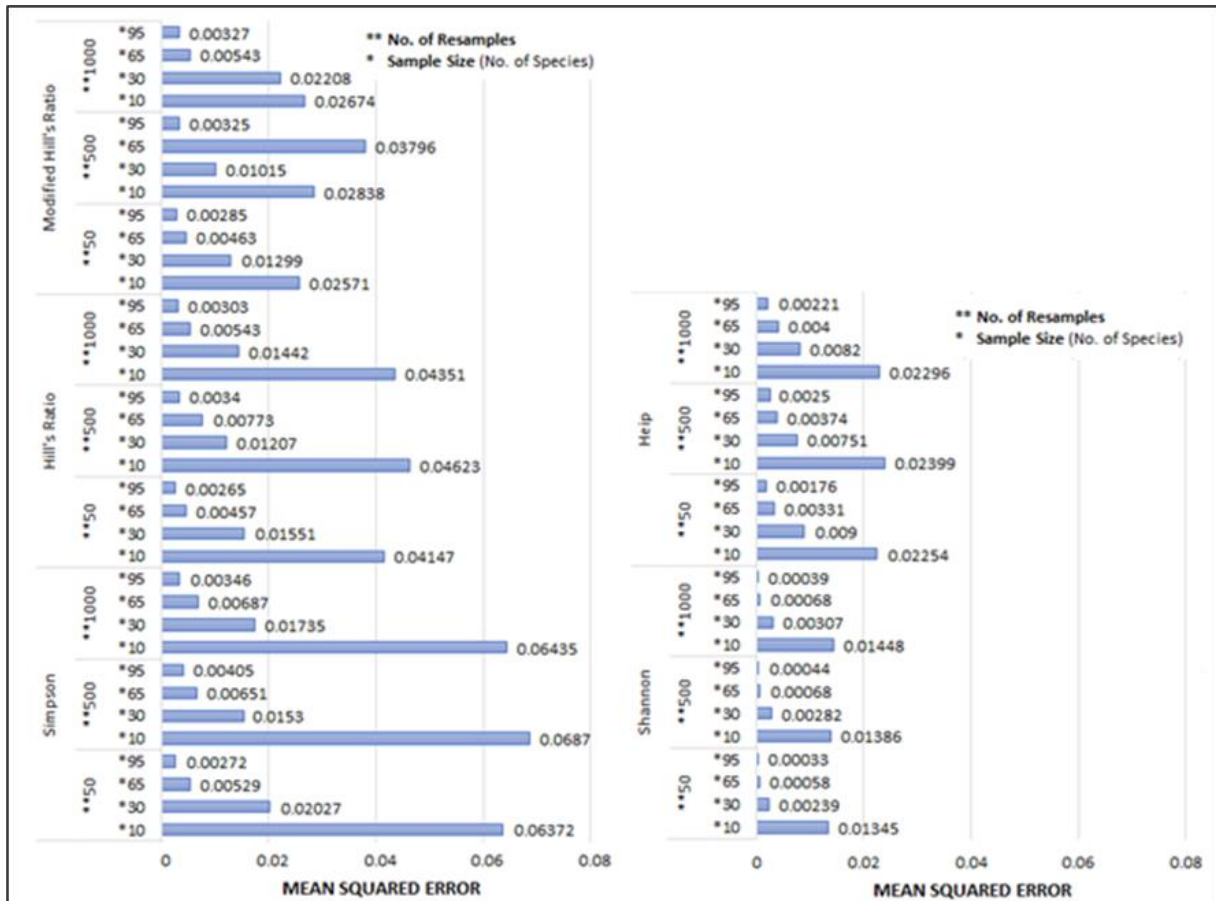

Figure 4. Bootstrap estimates for species evenness indices

On a different note, this paper also suggests that the Simpson diversity index provides more practical information and is appropriate for measuring the diversity of species of birds, since it produced a smaller degree of errors in the conducted simulation. According to Fontana et al. (2011), the Simpson diversity index emphasizes a community's evenness and being less sensitive to species richness. Their paper recorded the high impact of urbanization on the community composition of the avian community. Moreover, the Simpson index is preferred compared to the Shannon diversity index since the Shannon index is more sensitive to sample sizes. The minimum sample size needed to compute the Simpson diversity index is 15-20 (Williams et al., 2005).

Lastly, in measuring the evenness of birds' species, the Shannon evenness index generated the least errors compared to the other indices included in the simulation. This suggests that the Shannon evenness index is more efficient and will give more accurate estimates. According to Woldemariam (2018), the Shannon index is useful in measuring avifauna species' evenness. Besides, according to Redowan (2015), the Shannon evenness index accurately captured the right evenness of species, the spatial patterns they observed or the remotely sensed data agreed with the ground information.

Conclusion

In this simulation, all of the bootstrap sampling distribution of the identified indices for species richness, diversity, and evenness are typically distributed, as shown in Table A.1. The bootstrap estimates of these indices were generated and compared. Among the species richness indices, Menhinick's richness index is the most efficient given that it caused the least errors among the other indices. While among the diversity indices under the study, Simpson's diversity index produced the least MSE values indicating that it is the most efficient index. Moreover, Shannon's evenness index is the most efficient among the identified species evenness indices since it produced the least errors. For future simulation studies, one may consider different datasets to validate this study's findings. One may also account for other diversity indices such as Brillouin, Camargo evenness index, Fisher's Alpha, and one may also use other simulation procedures like Jackknife resampling to evaluate the index empirically.

References

Antoney, P.U. (1970). Bird communities in the forest habitats of Wayanad, South India. Shodhganga@INFLIBNET: Bird communities in the forest habitats of Wayanad, South India. https://shodhganga.inflibnet.ac.in/handle/10603/33672.

Brown, R., Jacobs, L.A., & Peet, R.K. (2007). Species Richness: Small Scale. *Encyclopedia of Life Sciences*. https://doi.org/10.1002/9780470015902.a0020488

Chivian, E., & Bernstein, A. (2010). How Our Health Depends on Biodiversity. Center for Health and the Global Environment.

Efron, B., & Tibshirani, R.J. (1993). An Introduction to the Bootstrap. Chapman & Hall.

Fisher, R. (1921). On the Mathematical Foundations of Theoretical Statistics. Royal Society.

Fontana, S., Sattler, T., Bontadina, F., & Moretti, M. (2011). How to Manage the Urban Green to Improve Birds Diversity and Community Structure. Lands and Urban Planning.

Gross, L., Beals, M., & Harrell, S. (2019). Diversity Indices: Shannon's H and E. *Quantitative Biology at Community Colleges, QUBES Educational Resources*. https://doi.org/10.25334/Q4844Q

Krebs C. (1999). Ecological Methodology. Benjamin Cummings, San Francisco, CA.

Launchbaugh, K. (2009). *Principle of Vegetation Measurement, Assessment, Ecological Monitoring, and Analysis*. Vegetation Measurement & Assessment. https://www.webpages.uidaho.edu/veg_measure/.

Nikulin, M. (1994). *Efficiency of a Statistical Procedure*. Encyclopedia of Mathematics, Springer.

Odum, P. (2000). *Fundamentals of Ecology*. Quezon City, Philippines: JMC Press, Inc.

Rao, C. (1962). Efficient Estimates and Optimum Inference Procedures in Large Samples. Journal of the Royal Statistical Society. Series B (Methodological), 24(1), 46-72. Retrieved February 12, 2020, from http://www.jstor.org/stable/2983745

Redowan, M. (2015). Spatial pattern of tree diversity and evenness across forest types in Majella National Park, Italy. *Forest Ecosystems*, *2*(24). https://doi.org/10.1186/s40663-015-0048-1

Trapani, L., & Horváth, L. (2016). Statistical inference in a random coefficient panel model. *Journal of Econometrics, Elsevier*, *193*(1), 54–75. https://doi.org/10.1016/j.jeconom.2016.01.006

Williams, V. L., Witkowski, E. T. F., & Balkwill, K. (2005). Application of diversity indices to appraise plant availability in the traditional medicinal markets of Johannesburg, South Africa. *Biodiversity and Conservation*, *14*, 2971–3001. https://doi.org/10.1007/s10531-004-0256-4

Woldemariam, W. M., Gutema, T.M., Morrison, K., & Aticho, A. (2018). Assessment of wetland flora and avifauna species diversity in Kafa Zone, Southern Ethiopia. *Journal of Asia-Pacific Biodiversity, 11*(4): 494-502. https://doi.org/10.1016/j.japb.2018.08.003