

Article

A Different Statistical Perspective on the Evaluation of Ecological Data Sets

Soner Yigit 

Biometry and Genetics Unit, Animal Science Department, Faculty of Agriculture, Canakkale Onsekiz Mart University, Canakkale 17020, Türkiye; soneryigit@comu.edu.tr

Abstract

Statistical significance varies depending on the sample size. Therefore, when the sample size is sufficient, even differences that affect the total variation very little may be statistically significant. For this reason, it is very important to report effect size measures that estimate the share of the difference between groups of samples in the total variation. This study aims to determine the most reliable effect size measures that can be used when evaluating data obtained from ecological studies. The three most popular effect size measures used in practice were compared in terms of their performance in 2700 different experimental conditions. For this purpose, random numbers generated from the multivariate Poisson distribution were used with the Monte Carlo simulation technique. As a result of the simulations, it was determined that Epsilon-squared and Omega-squared were quite unbiased estimators. Therefore, it was concluded that one of these two effect size measures should be reported in addition to the p -value when evaluating ecological studies.

Keywords: effect size; Epsilon-squared; Omega-squared; Eta-squared; statistical significance; sample size; practical significance

1. Introduction

In ecological studies, groups of samples such as sites, regions, stations, and locations are compared in terms of species abundance to gain information about community structure [1]. To test whether there is a statistically significant difference between such groups, several methods, such as Analysis of Similarity (ANOSIM) [2], Permutational Multivariate Analysis of Variance (PERMANOVA) [3], the Mantel test [4], Pillai's trace [5], and Permutational Analysis of Multivariate Dispersions (PERMDISP) [6], have been proposed. However, statistical significance is highly influenced by sample size; when the sample size is sufficiently large, even very small differences may become statistically significant [7–9]. This may lead to misleading, overinterpreted, or underinterpreted conclusions if statistical significance is considered in isolation. At this point, determining the share of the difference between the groups in the total variation is at least as important as hypothesis testing [10]. The share of groups in the total variation is called effect size, and the statistics that estimate it are called effect size measures [11]. However, it was observed that effect size measures were generally not taken into account when evaluating results in ecological studies. This may be partly because some of the statistical methods used (such as PERMANOVA and db-RDA), although they decompose the variation according to explanatory components, do not provide a summary statistic like ANOSIM which gives an R value that compares the mean similarity between groups to the mean similarity within groups. Although ANOSIM provides a coefficient, it does not show the share of the total variation explained by the



Academic Editor: Michael Wink

Received: 22 July 2025

Revised: 1 August 2025

Accepted: 4 August 2025

Published: 5 August 2025

Citation: Yigit, S. A Different Statistical Perspective on the Evaluation of Ecological Data Sets. *Diversity* **2025**, *17*, 555. <https://doi.org/10.3390/d17080555>

Copyright: © 2025 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

groups. In other words, it only tells whether the similarity between groups is greater than within groups, without quantifying the extent of this effect. This limitation highlights the need for separate effect size measures to determine how much variation can actually be attributed to group differences. Therefore, it is necessary to use effect size measures to estimate the share of groups in the total variation. A large number of effect size measures have been developed [10,12–17]. Ecological data generally do not meet the assumptions of ANOVA. Therefore, this was taken into account when selecting the effect size measures to be evaluated in the present study. Detailed simulation studies have shown that Epsilon-squared ($\hat{\epsilon}^2$) and Omega-squared ($\hat{\omega}^2$) are among the most reliable effect size measures even when the assumptions of classical ANOVA are violated [8,18,19]. Given that it is a fundamental and widely recognized measure of effect size, Eta-squared ($\hat{\eta}^2$) was also included in this study for reference purposes.

Accordingly, this study aims to evaluate the performance of these three effect size measures in ecological data sets and to identify the most reliable one(s) for ecological applications.

2. Materials and Methods

2.1. Effect Size Measures

In this study, $\hat{\eta}^2$, $\hat{\epsilon}^2$, and $\hat{\omega}^2$ were taken into consideration as effect size measures. In this section, the effect size measures in question are introduced.

$$\hat{\eta}^2 = \frac{SS_b}{SS_t} \quad (1)$$

$$\hat{\epsilon}^2 = \frac{SS_b - df_b \text{MSE}}{SS_t} \quad (2)$$

$$\hat{\omega}^2 = \frac{SS_b - df_b \text{MSE}}{SS_t + \text{MSE}} \quad (3)$$

where SS_b is the sum of squares between the groups, SS_t is the total sum of squares, df_b are the degrees of freedom between the groups, and MSE is the mean squared error [8,9]. The sums of squares, mean squares, and degrees of freedom in the above formulas are obtained from PERMANOVA [3].

2.2. Design of the Simulation Study

In this study, bias was used as the criterion for comparing effect sizes. For this purpose, random numbers generated according to the experimental conditions considered were used. These numbers were drawn from the multivariate Poisson distribution, which is suitable for modeling the count of events occurring within a specified time or area, to simulate species counts [20]. In the study, multivariate Poisson distributions with 75 different parameter combinations based on varying lambda (1, 3, 5, 7, and 10), number of variables ($p = 2, 3, 4, 5$, and 10), and correlation structures ($r = 0, 0.3$, and 0.6) were used. In order to create differences between groups, the same fixed numbers were added to all observations in one random group ($\delta = 0.3, 0.9$, and 1.5). Finally, the population effect size was estimated for each effect size measure in three different sample sizes ($n = 2, 3$, and 4) and four different group combinations ($k = 2, 3, 4$, and 10). This process was repeated 10,000 times and the means were calculated. The absolute difference between the mean of each effect size measure obtained as a result of 10,000 simulations and the true value of the population effect size was considered as bias. The population effect size was calculated as follows.

$$SS_b = N \sum_{i=1}^k \sum_{j=1}^p (\mu_{ij} - \mu_j)^2 \quad (4)$$

$$SS_w = N \sum_{i=1}^k \sum_{j=1}^p \sigma_{ij}^2 \quad (5)$$

$$\text{Population Effect Size} = \frac{SS_b}{SS_b + SS_w} \quad (6)$$

where k is the number of populations being compared, p is the number of variables (species), and σ^2 is the variance of each population with respect to each variable [11]. Simulation trials were conducted with codes written using the R programming language [21].

As a result of the simulation trials (10,000 simulations for each condition), predictions were made for 2700 different trial conditions for each test. This means a total of 8100 predictions for three tests. Unlike classical simulation studies, regression trees [22] were used to evaluate simulation results. When using the regression tree, the data was randomly divided into two groups: training (70% of all data) and testing (30% of all data). The model that creates the regression tree that is easiest to interpret, with the highest explained variation, was validated on test data. Thus, the factors affecting the predictions were determined more objectively.

3. Results

The variations explained by the regression trees obtained from the training and test data sets are 98.36% and 98.39%, respectively (Table 1). It can be seen that the variation explained and other model selection criteria by both trees are quite close. This shows that the interpretations made from the regression tree created for the test data are reliable.

Table 1. Model diagnostics for training and test regression trees.

Statistics	Training	Test
R-squared	98.36%	98.39%
Root-mean-squared error (RMSE)	1.7681	1.7553
Mean squared error (MSE)	3.1261	3.0811
Mean absolute deviation (MAD)	1.0819	1.0797
Mean absolute percent error (MAPE)	3.0629	2.6872

According to the regression tree results, the factor that most affects bias is the effect size measure used. However, the number of replications and the number of groups are as effective as 16% and 3.8% of the effect size measure, respectively. On the other hand, the shape of the distribution (mean), effect size (delta), number of variables (p), and correlation coefficient (r) between variables did not affect bias at all (Figure 1).

In terms of bias, effect size measures were divided into two groups (Figure 2). The first group included $\hat{\epsilon}^2$ and $\hat{\omega}^2$. $\hat{\epsilon}^2$ and $\hat{\omega}^2$ were not affected by any factors and their estimates deviated by a mean of 0.42% in all experimental conditions considered. This showed that $\hat{\epsilon}^2$ and $\hat{\omega}^2$ were quite unbiased estimators. $\hat{\eta}^2$, which has very high deviations in its estimates, was in the second group. Considering all experimental conditions, there was a mean deviation of 27.14% in $\hat{\eta}^2$'s estimates. The number of replications and the number of groups seriously affected the estimates of $\hat{\eta}^2$. For example, the mean deviations in the estimates were 38.88% when $n = 2$, 24.47% when $n = 3$, and 18.05% when $n = 4$. As in all experimental studies, increasing the number of replications gradually reduced bias. Moreover, the increase in the number of groups also negatively affected the bias of $\hat{\eta}^2$. For example, the bias was 34.58% when $n = 2$ and $k = 2$ and 3, and 43.29% when $n = 2$ and $k = 4$ and 10. Regardless of the experimental conditions considered, $\hat{\eta}^2$ made highly biased estimates relative to $\hat{\epsilon}^2$ and $\hat{\omega}^2$.

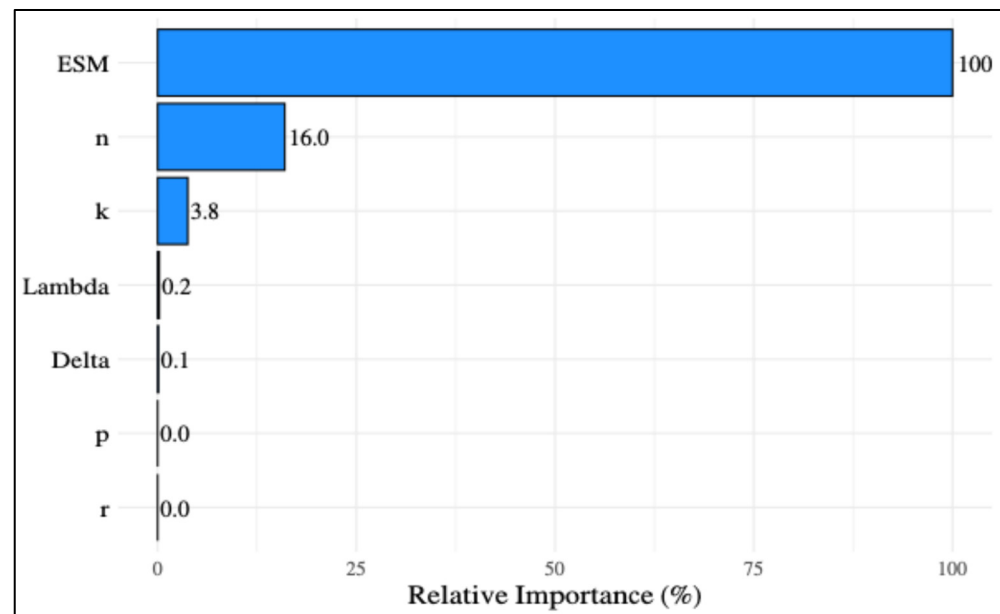


Figure 1. Relative importance of variables in the regression tree, where ESM is the effect size measure, n is the number of replications, k is the number of groups of samples, Lambda is the mean of Poisson distributions, Delta is the maximum difference between population means, p is the number of variables, and r is the correlation coefficient between variables.

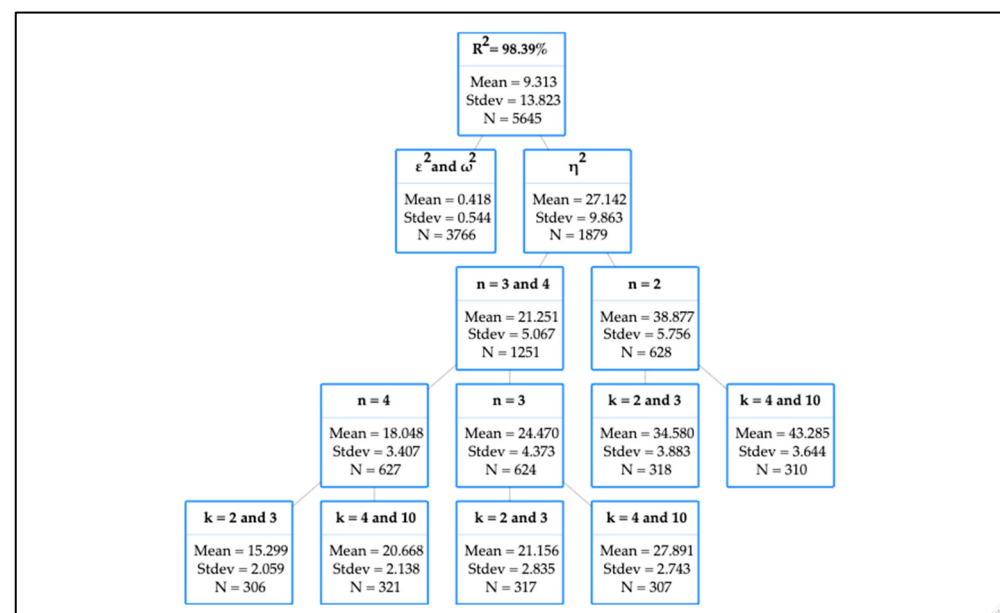


Figure 2. Regression tree for bias, where R^2 is the variation explained in the dependent variable (bias), Mean is the arithmetic average of bias, Stdev is the standard deviation of bias, N is the number of total observations, ϵ^2 is Epsilon-squared, ω^2 is Omega-squared, η^2 is Eta-squared, n is the number of replications, and k is the number of groups of sample.

4. Discussion

In general, the focus of this research is to investigate the sources of variation. With the widespread use of inferential statistics, the focus in evaluating study results has begun to shift. Results of scientific research have generally been evaluated on the basis of the p -value. As a result, the belief has emerged that as the p -value decreases, the results are considered more significant [23]. However, the p -value indicates the statistical significance of the difference observed between groups. In general, as the number of replications increases,

even very small differences begin to become statistically significant, due to the principle that the sample will more accurately represent the population. The reverse is also true. Significant differences observed between groups may not be statistically significant, due to the small number of replications [7–9]. To explain this situation more concretely, two different hypothetical examples are given in Appendix A. In the first of these examples, three groups were compared with respect to three variables using PERMANOVA, with four replications for each group. As a result of the analysis, it was seen that the differences between the groups were statistically significant ($p = 0.012$). However, differences between groups can explain 37.82% ($\hat{\epsilon}^2$) of the total variation. As can be noted, although the differences between the groups are statistically significant, they can only explain one third of the total variation. In the other hypothetical example, three groups were compared with two replications in terms of four variables. As a result of PERMANOVA, it was seen that there was no statistically significant difference between the groups ($p = 0.067$). On the other hand, it was determined that the statistically insignificant difference explained 73.54% of the total variation. This shows that there is a remarkable difference between the groups, but it is not statistically significant, due to the small number of replications. Therefore, evaluating results solely based on the p -value may be misleading. It is essential to report effect size measures alongside the p -value. Thus, reporting how much the differences between groups affect the total variation will eliminate contradictions and misunderstandings in the interpretation of the results [8–10,19,24]. However, the effect size measure you report should not be affected by sample size or any other factor. Therefore, the most appropriate effect size measures that can be used need to be determined. In this study, the three most commonly used effect size measures in practice were compared with a detailed simulation study based on ecological data sets. However, the estimates obtained empirically as a result of simulation trials were evaluated not subjectively, but with an objective method such as a regression tree. As in almost all studies conducted for classical ANOVA, it was observed that $\hat{\eta}^2$ gave quite biased results [8,18,19]. The bias of $\hat{\eta}^2$ gradually increased with the decrease in the number of replications. The shape of the distribution, effect size, number of variables, and correlation between variables generally did not affect the performance of effect size measures in any way. This situation is consistent with the results of other studies comparing effect sizes [8,18]. Regardless of the experimental conditions considered, $\hat{\epsilon}^2$ and $\hat{\omega}^2$ produced results that are quite unbiased compared to $\hat{\eta}^2$. Furthermore, the results of this study confirmed Glass and Hakstian [25] who claimed that the difference between the predictions of $\hat{\epsilon}^2$ and $\hat{\omega}^2$ is negligible. On the other hand, this study draws attention to the fact that when evaluating ecological data sets, one should not only focus on the p -value but also evaluate the share of the groups in the total variation. Therefore, it will contribute to making the results of ecological studies more understandable.

5. Conclusions

Particularly, the main goal of ecological studies comparing groups in terms of species abundance should be the separation of the total variation into its components. The p -value does not partition variation; it only indicates statistical significance, as demonstrated by two concrete examples. The most appropriate effect size measures that can be used for this purpose were determined by a detailed simulation study based on ecological data sets. As a result of the simulation study, it was revealed that $\hat{\epsilon}^2$ and $\hat{\omega}^2$ are quite unbiased estimators. It is thought that determining the sources of variation observed in terms of the subject of interest will make significant contributions to the studies conducted. For this, it is recommended to report $\hat{\epsilon}^2$ or $\hat{\omega}^2$ together with the p -value.

This study primarily focused on ecological data sets under certain assumptions, such as multidimensional Poisson distributions and relatively small sample sizes. While the

findings provide valuable insights, future research could examine the performance of effect size measures under different distributional scenarios and with larger, more diverse ecological data sets. Such work would help reinforce and broaden the applicability of these estimators across various ecological contexts.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Datasets are available from the corresponding author upon reasonable request.

Conflicts of Interest: The author declares no conflicts of interest.

Appendix A

Appendix A.1 Example 1

Table A1. Hypothetical data for first example.

	p1	p2	p3
g1	1	3	0
	5	4	2
	4	2	3
	2	0	6
g2	1	3	1
	1	3	1
	3	4	0
	2	4	3
g3	2	3	6
	10	2	7
	12	10	2
	11	4	6

Table A2. Results of PERMANOVA for first example.

Source	DF	SS	MS	F	p
Between Groups	2	156.67	78.34	4.34	0.012
Within Groups	9	162.25	18.03		
Total	11	318.92			

DF: degrees of freedom; SS: sum of squares; MS: mean square; F: F-value; p: probability value.

Table A3. Effect size estimates for first example.

η^2	ϵ^2	ω^2
0.4912464	0.3781901	0.3579555

ϵ^2 : Epsilon-squared; ω^2 : Omega-squared; η^2 : Eta-squared.

Appendix A.2 Example 2

Table A4. Hypothetical data for second example.

	p1	p2	p3
g1	1	2	5
	3	3	5

Table A4. *Cont.*

	p1	p2	p3
δ_2	0	1	1
	1	5	3
δ_3	28	2	5
	15	6	4

Table A5. Results of PERMANOVA for second example.

Source	DF	SS	MS	F	p
Between Groups	2	561.67	280.84	7.94	0.06667
Within Groups	3	106.00	35.33		
Total	5	667.67			

DF: degrees of freedom; SS: sum of squares; MS: mean square; F: F-value; p: probability value.

Table A6. Results of PERMANOVA for second example.

η^2	ϵ^2	ω^2
0.8412381	0.7353969	0.6984353

ϵ^2 : Epsilon-squared; ω^2 : Omega-squared; η^2 : Eta-squared.

References

- Anderson, M.J.; Walsh, D.C. Permanova, Anosim, and the Mantel test in the face of heterogeneous dispersions: What null hypothesis are you testing? *Ecol. Monogr.* **2013**, *83*, 557–574. [\[CrossRef\]](#)
- Clarke, K.R. Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* **1993**, *18*, 117–143. [\[CrossRef\]](#)
- Anderson, M.J. A new method for non-parametric multivariate analysis of variance. *Aust. J. Ecol.* **2001**, *26*, 32–46.
- Mantel, N. The detection of disease clustering and a generalized regression approach. *Cancer Res.* **1967**, *27*, 209–220. [\[PubMed\]](#)
- Pillai, K.C.S. Some new test criteria in multivariate analysis. *Ann. Math. Stat.* **1955**, *26*, 117–121. [\[CrossRef\]](#)
- Anderson, M.J. Distance-based tests for homogeneity of multivariate dispersions. *Biometrics* **2006**, *62*, 245–253. [\[CrossRef\]](#) [\[PubMed\]](#)
- Fan, X. Statistical significance and effect size in education research: Two sides of a coin. *J. Educ. Res.* **2001**, *94*, 275–282. [\[CrossRef\]](#)
- Yigit, S.; Mendes, M. Which effect size measure is appropriate for one-way and two-way ANOVA models?: A Monte Carlo simulation study. *REVSTAT-Stat. J.* **2018**, *16*, 295–313.
- Yigit, S. Comparison of some effect size measures in simple and multiple linear regression models. *Eskişehir Tech. Univ. J. Sci. Technol. A-Appl. Sci. Eng.* **2021**, *22*, 77–84. [\[CrossRef\]](#)
- Hays, W. *Statistics for Psychologists*, 1st ed.; Holt, Rinehart and Winston: New York, NY, USA, 1963.
- Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; Lawrence Erlbaum Associates: Mahwah, NJ, USA, 1988.
- Kelley, T.L. An unbiased correlation ratio measure. *Proc. Natl. Acad. Sci. USA* **1935**, *21*, 554–559. [\[CrossRef\]](#) [\[PubMed\]](#)
- Maxwell, S.E.; Camp, C.J.; Arvey, R.D. Measures of strength of association: A comparative examination. *J. Appl. Psychol.* **1981**, *66*, 525–534. [\[CrossRef\]](#)
- Keppel, G. *Design and Analysis A Researcher's Handbook*, 2nd ed.; Prentice Hall: Mahwah, NJ, USA, 1982.
- Olejnik, S.; Algina, J. Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemp. Educ. Psychol.* **2000**, *25*, 241–286. [\[CrossRef\]](#) [\[PubMed\]](#)
- Kirk, R.E. The importance of effect magnitude. In *Handbook of Research Methods in Experimental Psychology*, 1st ed.; Davis, S.F., Ed.; Blackwell: Oxford, UK, 2003; pp. 83–105.
- Grissom, R.; Kim, J. *Effect Sizes for Research: A Broad Practical Approach*, 1st ed.; Lawrence Erlbaum Associates: Mahwah, NJ, USA, 2005.
- Skidmore, S.T.; Thompson, B. Bias and precision of some classical ANOVA effect sizes when assumptions are violated. *Behav. Res. Methods* **2013**, *45*, 536–546. [\[CrossRef\]](#) [\[PubMed\]](#)
- Okada, K. Is omega squared less biased? A comparison of three major effect size indices in one-way ANOVA. *Behaviormetrika* **2013**, *40*, 129–147. [\[CrossRef\]](#)
- Yahav, I.; Shmueli, G. On generating multivariate Poisson data in management science applications. *Appl. Stoch. Models Bus. Ind.* **2012**, *28*, 91–102. [\[CrossRef\]](#)
- R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria. Available online: <https://www.R-project.org/> (accessed on 21 July 2025).

22. Breiman, L.; Friedman, J.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*, 1st ed.; Chapman and Hall/CRC: New York, NY, USA, 1984.
23. Nickerson, R.S. Null hypothesis significance testing: A review of an old and continuing controversy. *Psychol. Methods* **2000**, *5*, 241–301. [[CrossRef](#)] [[PubMed](#)]
24. Cumming, G.; Finch, S. A Primer on the Understanding, Use, and Calculation of Confidence Intervals that are Based on Central and Noncentral Distributions. *Educ. Psychol. Meas.* **2001**, *61*, 532–574. [[CrossRef](#)]
25. Glass, G.V.; Hakstian, A.R. Measures of Association in Comparative Experiments: Their Development and Interpretation. *Am. Educ. Res. J.* **1969**, *6*, 403–414. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.