

Bias Correction for Eta Squared in One-Way ANOVA

Xiaofeng Steven Liu¹ 

[1] *Department of Educational Studies, University of South Carolina, Columbia, SC, USA.*

Methodology, 2022, Vol. 18(1), 44–57, <https://doi.org/10.5964/meth.7745>

Received: 2021-10-28 • Accepted: 2022-02-21 • Published (VoR): 2022-03-31

Corresponding Author: Xiaofeng Steven Liu, Department of Educational Studies, University of South Carolina, Columbia, SC 29208, USA. +1-803-777-6084. +1-803-777-7741. E-mail: xliu@email.sc.edu

Abstract

Eta squared is a popular effect size, but contains positive bias. Bootstrapping can be used to remove the bias from eta squared. Compared to epsilon squared and omega squared, bootstrap bias correction does not make distributional assumption, and it is easy to implement. A real example and computer simulations are included to illustrate its application. The bootstrap bias-corrected eta squared shows very little bias, and it serves as a good alternative to eta squared and epsilon squared, the latter of which can turn negative in some circumstances.

Keywords

eta squared, bias, bootstrap

Eta squared (η^2) is one of the most common effect sizes in the social sciences. It has a long history dating back to [Pearson \(1923\)](#) and [Fisher \(1925\)](#), and it occupies a central place in relation to other effect sizes. For instance, it is equivalent to the coefficient of determination in regression analysis, and the square root of eta squared is the correlation between the predicted and actual outcomes. Also, it is functionally related to Cohen's effect size (f), which plays an important role in statistical power analysis. Cohen's effect size which represents the standard deviation of standardized group means can be easily computed from eta squared.

$$f = \sqrt{\frac{\eta^2}{1 - \eta^2}} \quad (1)$$

Eta squared will remain a staple effect size measure in publications because journals have been increasingly encouraging researchers to include effect size measures. American Psychological Association recommended effect size measures in addition to statistical



tests (Wilkinson & Task Force on Statistical Inference, 1999). American Educational Research Association (2006) also adopted a similar policy on effect size reporting.

Despite its popularity, eta squared is not an unbiased estimate and is known to have positive bias. The existence of the bias can be deduced from the computation of eta squared.

$$\eta^2 = \frac{SSB}{TSS} \quad (2)$$

where *SSB* is the between-group sum of squares and *TSS* is the total sum of squares. The two sums of squares are unbiased estimates on their own, but their ratio is not necessarily an unbiased estimate. The bias of eta squared is well documented in the literature (Keselman, 1975; Okada, 2013). The bias depends on the sample size and the population effect size, and it can be substantial in some cases.

The bias of eta squared has not really diminished its popularity because of its straightforward interpretation (Mordkoff, 2019). In the context of one-way ANOVA, eta squared can be interpreted as the proportion of the variation in the outcome associated with the group membership (i.e., treatment factor). Its alternatives (i.e., epsilon squared and omega squared) have not completely unseated the dominant role of eta squared in social science research. Matter of fact, eta squared is routinely reported in journal publications and is readily accessible in popular statistical software (e.g., SPSS).

The two alternative effect size measures (ϵ^2 and ω^2) are created to lessen the bias of eta squared. The epsilon squared includes degrees of freedom to correct potential bias in estimating the population effect size (Kelley, 1935).

$$\epsilon^2 = 1 - \frac{(N - 1)SSE}{(N - g)TSS} \quad (3)$$

where *N* is the total sample size, *g* is the number of treatment groups, and *SSE* is the within-group sum of squares in ANOVA. It is easy to see that $SSE/(N - g)$ is an unbiased estimate of the within group variance, and that $TSS/(N - 1)$ is an unbiased estimate of the total variance. Epsilon squared reduces the positive bias in eta squared, but can result in underestimation. In a similar vein, Hays (1963) derived omega squared to decrease the positive bias in eta squared.

$$\omega^2 = \frac{SSB - (g - 1)MSE}{TSS + MSE} \quad (4)$$

where *MSE* is the within-group means squares. Although ω^2 and ϵ^2 have less positive bias than η^2 , they are not unbiased and can underestimate the population effect size. The three effect size estimates generally follow a descending order, $\eta^2 \geq \epsilon^2 \geq \omega^2$. It is recommended that ϵ^2 be used in lieu of η^2 (Okada, 2013). It appears to be a compromise

choice among the three competing effect size measures because the value of ϵ^2 sits between η^2 and ω^2 . However, neither ϵ^2 nor ω^2 has gained the same popularity as η^2 in practice.

So far, there is no unbiased estimate of the population effect size, as the less biased alternatives (ϵ^2 and ω^2) are not unbiased and can underestimate the population effect size. Both epsilon squared and omega squared implicitly use normality assumption in correcting the bias in eta squared (see Kelley, 1935; Hays, 1963). However, the normality assumption may not always be tenable in many practical situations, where limited data in a study do not lend support to the normality assumption. The actual data distribution can be difficult to ascertain, due to scarcity of the data in a small study. It will be advisable to remove the bias in eta squared without making distributional assumption. Non-parametric bootstrapping becomes a very viable way to correct the bias in eta squared in view of the increasing computing capability of modern computers.

Bootstrap Bias Correction

Bootstrapping can be used to estimate bias, although it is often deployed to estimate a complex statistic, which is analytically intractable. Bootstrapping draws repeated samples from the original sample data, which are conceived of as a proxy population. The repeated samples from this proxy population are used to imitate the sampling distribution of the statistic in question. The statistic can be estimated from the sampling distribution of the bootstrap samples. The inference from the sample to the population is then made analogous to the inference from the bootstrap sample to the original sample. For instance, we can use bootstrapping to estimate the standard error of a ratio. As the standard error of the ratio is analytically difficult to derive, bootstrapping can be utilized to compute the standard error of the ratio. The original sample data serves as a proxy population. We repeatedly draw simple random samples with replacement from the original data or the proxy population. Those repeated samples are bootstrap samples, for which the ratios are computed. The ratios from the bootstrap samples form a sampling distribution, from which the standard error can be obtained. In theory, bootstrapping can be used to calculate bias. As the bias estimate is just a statistic like the standard error, bootstrapping applies equally well in this situation (Chapter 10, Efron & Tibshirani, 1998).

Efron and Tibshirani (1998) provided the statistical theory behind bootstrap bias correction. It is relatively easy to empirically test whether bootstrap bias estimation works by way of Bessel's correction, which uses the adjusted degrees of freedom in calculating the sample variance. If the sum of squared deviations, SS , were divided by the sample size, n , the sample variance thus computed would have a negative bias of $-\sigma^2/n$. Bootstrapping can be used to correctly estimate the negative bias in SS/n . We leave out the code for brevity of the presentation, but it just shows the veracity of bootstrap bias

estimation in the case of a well-known bias. The flexibility of bootstrap bias estimation can be further attested by the positive bias in eta squared.

The bias of an estimator is generally defined as the difference between the expectation of the parameter estimator $\hat{\theta}$ and the parameter itself θ .

$$\text{bias} = E(\hat{\theta}) - \theta \quad (5)$$

If the bias is positive, it means that the estimator on average is larger than the population parameter. If the bias is negative, it means that the estimator on average tends to underestimate the population parameter. When the bias is zero, the estimator is said to be an unbiased estimator. Knowing the bias, we can correct an estimator and make it unbiased. The bias corrected estimator is, therefore, $\hat{\theta} - \text{bias}$.

The bootstrap estimate of bias ($\widehat{\text{bias}}$) is the expectation of the estimators of the bootstrap samples ($\hat{\theta}^*$) minus the original estimator ($\hat{\theta}$):

$$\widehat{\text{bias}} = E(\hat{\theta}^*) - \hat{\theta} \quad (6)$$

To understand the bootstrap estimate of bias, $\widehat{\text{bias}}$, we can analogize $E(\hat{\theta}^*)$ to $E(\hat{\theta})$ and $\hat{\theta}$ to θ in the definition of bias. First, $\hat{\theta}^*$ is called the bootstrap replicate. It is the relevant statistic based on a bootstrap sample, but it is computed in the same way as the estimator $\hat{\theta}$ in the original sample. We repeatedly generate bootstrap samples of the same size as the original sample. For each bootstrap sample, we calculate the bootstrap replicate, $\hat{\theta}^*$. Together, the bootstrap replicates or $\hat{\theta}^*$ s form a sampling distribution, which is analogous to the sampling distribution of $\hat{\theta}$. Thus, $E(\hat{\theta}^*)$ is made comparable to $E(\hat{\theta})$. The expectation of the bootstrap replicate $\hat{\theta}^*$ can be readily computed as the mean of all the bootstrap replicates, based on B number of bootstrap samples.

$$E(\hat{\theta}_b^*) = \left(\frac{1}{B} \sum_{b=1}^B \hat{\theta}_b^* \right) \quad (7)$$

where $\hat{\theta}_b^*$ is a bootstrap replicate of the original estimator, based on the b th bootstrap sample. Second, the bootstrap theory suggests that the original sample data is the proxy population, from which all bootstrap samples are drawn. In other words, $\hat{\theta}$ is the proxy population parameter, relative to all the bootstrap replicates, $\hat{\theta}_b^*$ s. Thus, $\hat{\theta}$ is made analogous to θ . We can then compute the bootstrap estimate of bias as

$$\begin{aligned} \widehat{\text{bias}} &= E(\hat{\theta}_b^*) - \hat{\theta} \\ &\cong \left(\frac{1}{B} \sum_{b=1}^B \hat{\theta}_b^* \right) - \hat{\theta} \end{aligned} \quad (8)$$

The last expression in the equation is the bootstrap estimate of bias, $\widehat{\text{bias}}$.

In our case, the statistic is the eta squared, η^2 , and the parameter is the population eta squared. Therefore, the bootstrap bias estimate for η^2 is

$$\widehat{bias} \cong \left(\frac{1}{B} \sum_{b=1}^B \eta_b^{2*} \right) - \eta^2 \quad (9)$$

where η_b^{2*} is the replicate statistic from the bootstrap sample. The bias corrected eta squared, η_c^2 , then becomes twice the original estimator minus the mean of bootstrap replicates:

$$\begin{aligned} \eta_c^2 &= \eta^2 - \widehat{bias} \\ &= 2\eta^2 - \left(\frac{1}{B} \sum_{b=1}^B \eta_b^{2*} \right) \end{aligned} \quad (10)$$

The bias estimator converges to a certain limit as the number of bootstrap samples, B , increases to infinity. In practice, we do not need to obtain an infinite number of bootstrap samples and replicates. The number of bootstrap samples, B , can be increased gradually until the mean of bootstrap replicates shows convergence to a certain value. The means of bootstrap replicates can be lined up against the increasing number of bootstrap replicates, B . The limit of convergence can be easily identified on a graph (Efron, 1982; Efron, 1990; Efron & Tibshirani, 1998).

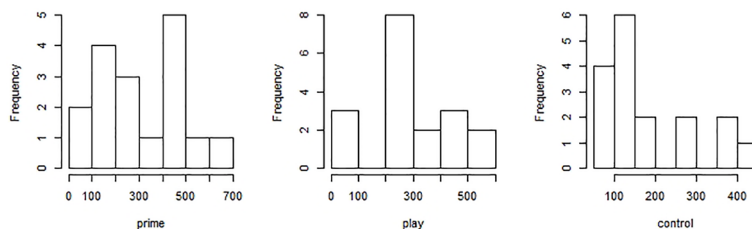
Example

The example data come from a randomized comparative experiment on the effects of money on human behavior (Vohs, Mead, & Goode, 2006). Money was believed to change human motivation and behavior. When people were mentally reminded of the idea of money (a priming technique), they would act more self-sufficiently and would be less likely to request for help. In the experiment, subjects were randomly assigned to three conditions. They were reminded of money (prime condition) or play money (play condition) or not reminded of money (control condition) while they were asked to descramble several words and create sensible phrases using four out of five jumbled words. It was hypothesized that participants who were reminded of money and play money would work longer than participants in the control before requesting for help. The outcome was time in seconds before asking for help. Table 1 lists the data (see Moore, Notz, & Fligner, 2015, p. 658). The ANOVA analysis confirmed the research hypothesis ($F = 3.73$, $p = .031$).

Table 1*Time in Seconds Until Subjects Ask for Help*

Condition	Time in seconds										
prime	609	444	242	199	174	55	251	466	443	531	135
	241	476	482	362	69	160					
play	455	100	238	243	500	570	231	380	222	71	232
	219	320	261	290	495	600	67				
control	118	272	413	291	140	104	55	189	126	400	92
	64	88	142	141	373	156					

The eta squared is $\eta^2 = .1321$. Despite of being a small fractional number, it is very close to Cohen's rule of thumb number for large eta squared .1379 (Cohen, 1988, p. 287). It should be noted that Cohen's rule of thumb numbers are used here for reference, and that the rule of thumb numbers do not dictate the magnitude of effect size in a specific field. The eta squared is known to have positive bias so that the true effect size is definitely smaller. The less biased epsilon squared is $\epsilon^2 = .0966$. According to the epsilon squared, the bias is approximately $\eta^2 - \epsilon^2 \approx .0354$. Since all these numbers are small fractionals, it makes sense to see the size of the bias through the percentage of the effect size measure. The percentage of the bias in η^2 is $(\eta^2 - \epsilon^2)/\eta^2 = .268$ or 26.8%. So, there is a sizable bias in the estimated eta squared.

Figure 1*Histograms of Time Among the Three Groups*

The epsilon squared is less biased, but is not unbiased. It is implicitly based on the normality assumption. Without the assumption of data distribution, it is not feasible to find the expectations of those moment estimates (i.e., sums of squares), from which the epsilon squared is derived. A simple analysis of the data in this experiment suggests that the data are not normally distributed (see Figure 1). The histograms of the data in the three conditions basically confirms that. As ANOVA is fairly robust against the violation

of the normality assumption, it can be validly applied. Violation of the distribution assumption, however, affects the accuracy of epsilon squared. Given the limited data, it is difficult, if not impossible, to ascertain the true distribution of the data and, in turn, the expectations of the sum of squares in the effect size measures, be it η^2 or ϵ^2 .

Non-parametric bootstrapping becomes a very practical tool to estimate the bias in this situation. The technique does not depend on distributional assumption. The bootstrap bias estimate converges to .0320 as the number of bootstrap replicates gradually increases (see R code in the [Appendix](#)). By the law of large numbers, the mean averaged over a large number of replicates will approach the expectation. The converged bias estimate can be easily identified in [Table 2](#) and [Figure 2](#). Some bias estimates are above the converged value, whereas other bias estimates are below the converged value (e.g., $B = 100$ and $B = 200$).

Table 2

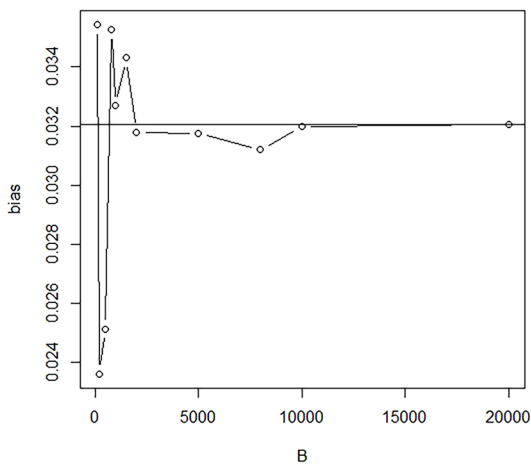
Bias Estimates and the Number of Bootstrap Replicates

<i>B</i>	100	200	500	800	1000	1500	2000	5000	8000	10000	20000
Bias estimate	.0354	.0236	.0251	.0353	.0327	.0343	.0318	.0318	.0312	.0320	.0320

Note. *B* = Number of bootstrap replicates.

Figure 2

Convergence of Bias Estimates



[Figure 2](#) is similar to a typical graph that illustrates the convergence in the law of large numbers. Some numbers are left out on the horizontal axis to show

the pace of convergence. The bootstrap bias-corrected eta squared is, therefore, $\eta_c^2 = \eta^2 - \widehat{bias} = .1321 - .0320 = .1001$. Compared with epsilon squared 0.0966, the bootstrap bias-corrected eta square is slightly larger, that is, $\eta_c^2 > \epsilon^2$. The epsilon squared implies a bias estimate of .0354, whereas the bootstrap bias estimate is .0320 or 9.6% smaller. It makes sense to obtain a smaller bootstrap bias estimate because ϵ^2 is known to over correct the positive bias and underestimate the true effect size.

Compared to bootstrapping, bias correction through epsilon squared depends more on the size of eta squared and sample sizes. As a result, epsilon squared can easily return a negative estimate. Simple algebra illustrates this point. Epsilon squared can be expressed in terms of eta squared (Mordkoff, 2019, eq. 5).

$$\epsilon^2 = \eta^2 - \frac{g-1}{N-g}(1-\eta^2) \quad (11)$$

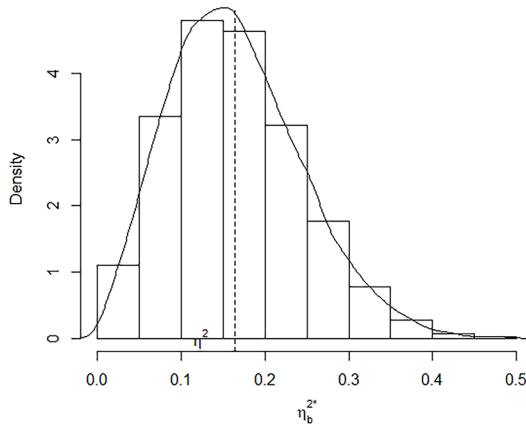
The bias estimate implied by ϵ^2 is $(g-1)/(N-g)(1-\eta^2)$. The bias correction is always positive but can exceed η^2 . When eta squared is smaller than the ratio of $(g-1)/(N-1)$, epsilon squared is negative, that is, $\eta^2 < (g-1)/(N-1)$ means $\epsilon^2 < 0$. For instance, in a one-way ANOVA design with three groups of equal sample size 10, we have $g = 3$, $N = 30$, and $(g-1)/(N-1) = .069$. If η^2 is less than .069, then ϵ^2 is negative. It is an overestimate of potential bias in η^2 because the population effect size cannot be negative. Since η^2 of .069 is more than Cohen's rule of thumb number .0588 for the medium-sized η^2 , ϵ^2 will be negative for all the small and medium-sized η^2 with the limited sample sizes (i.e., $g = 3$ and $N = 30$). Even if the total sample size N increases to 60, ϵ^2 will still be negative for Cohen's small-sized eta squared (i.e., $\eta^2 \leq .0099$). In other words, ϵ^2 will likely result in a negative estimate in many situations where eta squared or sample size is not large. This may partially explain why η^2 is still a popular effect size measure: it is never negative in contrast to its less biased alternatives, ϵ^2 and ω^2 . Bootstrap bias-corrected estimate less likely suffers the same shortcoming because the accuracy of bootstrapping in theory depends less on the magnitude of the effect size and sample sizes but more on the number of bootstrap replicates.

The sampling distribution of eta squared can be approximated by bootstrap replicates of eta squared, which provides clues of the potential bias. The bootstrap procedure shows that the sampling distribution of the eta squared is right skewed in the example: its bias may not solely depend on the sample size. The original eta squared measure of .1321 is marked on the sampling distribution in Figure 3, and it has a quantile score of .37 in the sampling distribution. In other words, its bootstrap replicates, η_b^{2*} , exceed the original η^2 63% of the time among repeated bootstrap samples – a clear positive bias. The expectation of the bootstrap replicates or $E(\eta_b^{2*})$ can be computed by averaging 20,000 bootstrap replicates, and the average is .16455, marked by the dotted line in Figure 3. There does not appear a ready formula to describe the sampling distribution or behavior of eta squared in repeated samples. The bias in eta squared may be too complex to yield

a simple analytical solution. Thus, epsilon squared can lessen the bias of the eta squared by using sample sizes, but cannot eradicate the bias. Bootstrapping appears to be a viable alternative.

Figure 3

Bootstrap Sampling Distribution of η_b^{2}*



Simulations

Simulations can be used to show that the bootstrap bias-corrected eta squared generally performs better than eta squared and epsilon squared under the normality assumption. When data show a skewed distribution or mix normal distribution, the bootstrap bias-corrected eta squared does not show much bias either. [Table 3](#) uses the same parameter setting for effect sizes as in [Okada \(2013\)](#). Three effect sizes are .26471 (large), .12339 (medium), and .02200 (small). The R code for the simulation is adapted from [Okada \(2013\)](#). A one-way ANOVA with four groups is used with varying group size (n). The sample size (n) goes from small (5) to large (30). As bootstrapping is a computation intensive method, the number of sample datasets and the number of bootstrap replicates are limited to thirty and one thousand, respectively. The bias estimate is the average difference between the effect size and its estimate. [Table 3](#) lists the bias estimates for eta squared, epsilon squared, and the bootstrap bias-corrected eta squared. The simulation duplicates the finding about eta squared and epsilon squared in [Okada \(2013\)](#). The eta squared is positively biased, and the epsilon squared is negatively biased. The bias diminishes with increased sample sizes. The bootstrap bias-corrected eta squared generally shows little bias, compared to eta squared and epsilon squared.

Table 3*Bias Estimates for η^2 , ϵ^2 , and η_c^2 Under Normality*

Effect size	<i>n</i>	η^2	ϵ^2	η_c^2
.26471	5	.07464	-.04924	-.02687
	10	.05239	-.00452	.00214
	20	.02095	-.00725	-.00483
	30	.01748	-.00108	-.00034
.12329	5	.09979	-.04589	-.02015
	10	.06305	-.00475	.00338
	20	.02903	-.00443	-.00203
	30	.02133	-.00079	.00004
.02200	5	.13340	-.02497	.00229
	10	.07037	-.00527	.00309
	20	.03628	-.00089	.00100
	30	.02435	-.00031	.00033

The second simulation in [Table 4](#) shows that the bootstrap bias-corrected eta squared does not have much bias when the data follow a skewed or mix normal distribution. The simulation uses a gamma distribution for the skewed distribution. The gamma distribution can show a variety of distribution shapes. The shape parameter is set at 3, and the rate parameter is set to $\sqrt{3}$. The gamma distribution is then centered at $\sqrt{3}$, and the resulting distribution has a zero mean and unit standard deviation. A similar strategy is used to generate the mix normal distribution. The mix normal distribution has two means (0 and 2) and two standard deviations (1 and 1) with a mix proportion (.5 and .5). It is then standardized to have a zero mean and unit standard deviation. As non-normal distribution poses its own challenge in ANOVA and effect size interpretation, researchers may choose to transform the original data. For instance, researchers often apply log transformation to skewed distributions to make data look normal. It begs the question whether eta squared should be used in this case. So, the simulation is limited to the special case of zero effect. In other words, it is more relevant to see whether the bootstrap bias-corrected eta squared can rightfully identify no effect under the background noise of non-normal distributions. The results in [Table 4](#) confirms that the bootstrap bias-corrected eta squared shows very little bias.

Table 4*Bias Estimates for η^2 and η_c^2 Under Non-Normal Distributions*

Distribution	<i>n</i>	η^2	η_c^2
skewed	5	.15667	.02687
	10	.06877	-.00121
	20	.04898	.01323
	30	.02882	.00464
mix normal	5	.13642	.00128
	10	.09315	.02316
	20	.04790	.01182
	30	.02449	.00003

Conclusion

The effect size measure η^2 may contain sizable bias in one-way ANOVA. Given the cost of running a scientific study, it is advisable to remove the bias from the η^2 and offer a bias-corrected estimate. Bootstrapping is a very economical way to calculate the bias estimate because it only involves a little coding effort and computing time. The bootstrap bias-corrected eta squared can show substantial improvement as demonstrated in the example. Compared to less biased effect size measure ϵ^2 , the bootstrap bias estimate appears slightly better. In the example the bootstrap bias estimate is rightly smaller than the bias implied by the epsilon squared, which can overestimate the bias in eta squared.

The advantage of bootstrap bias estimation is no prior knowledge about data distribution. When data are limited as in many situations, they do not usually appear normal. It is often difficult to ascertain the actual distribution, based on the limited data. Nevertheless, analysis of variance is still applicable, due to its robustness, and eta squared is often reported. Non-parametric bootstrapping can be used to estimate the bias in eta squared without prior knowledge of the data distribution. The bias of the eta squared in ANOVA can potentially arise from the estimation method and data distribution. Non-parametric bootstrap offers a nice solution to correct the bias arising from both sources. In bootstrapping the bias is allowed to have a complex relation to the contributing factors (e.g., sample size, data distribution, etc.).

Bootstrap bias estimate is also a good alternative when ϵ^2 is negative. When η^2 or sample size is not large, ϵ^2 can be negative. In practice, a small or medium-sized η^2 is more likely than a large η^2 . The same thing can be said about sample size. So, negative ϵ^2 is likely in some cases. Also, bootstrapping can be used to show the long-run behavior of η^2 estimates. Its sampling distribution can be readily obtained from the bootstrap replicates, which can graphically illuminate the size of the potential bias.

Finally, it should be noted that eta squared is a small number. Its bias estimate is often an even smaller number. There are circumstances when eta squared and its bias may not convey a difference of practical importance. Therefore, its use requires nuanced interpretation with reference to the context and the intended audience. Other effect size measures can also be considered: probability of superiority (Ruscio & Gera, 2013), common language effect size (Brooks, Dalal, & Nolan, 2014; McGraw & Wong, 1992; Li, Nesca, Waisman, Cheng, & Tze, 2021; Li & Tze, 2021), and binomial effect size. Although ANOVA is a fixed procedure, a variety of effect size measures can be used and modified to meet different needs. Future studies in this area can focus on effect size measures in non-normal data situations, where a regular effect size such as eta squared needs to be adapted to allow for meaningful interpretation.

Funding: The author has no funding to report.

Acknowledgments: The author has no additional (i.e., non-financial) support to report.

Competing Interests: The author has declared that no competing interests exist.

References

- American Educational Research Association. (2006). Standards for reporting on empirical social science research in AERA publications. *Educational Researcher*, 35(6), 33–40. <https://doi.org/10.3102/0013189X035006033>
- Brooks, M. E., Dalal, D. K., & Nolan, K. P. (2014). Are common language effect sizes easier to understand than traditional effect sizes? *The Journal of Applied Psychology*, 99(2), 332–340. <https://doi.org/10.1037/a0034745>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ, USA: Erlbaum.
- Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics.
- Efron, B. (1990). More efficient bootstrap computations. *Journal of the American Statistical Association*, 85(409), 79–89. <https://doi.org/10.1080/01621459.1990.10475309>
- Efron, B., & Tibshirani, R. J. (1998). *An introduction to the bootstrap*. Boca Raton, FL, USA: CRC Press LLC.
- Fisher, R. A. (1925). *Statistical methods for research workers*. Edinburgh, United Kingdom: Oliver & Boyd.
- Hays, W. L. (1963). *Statistics*. New York, NY, USA: Holt, Rinehart, & Winston.
- Kelley, T. L. (1935). An unbiased correlation ratio measure. *Proceedings of the National Academy of Sciences of the United States of America*, 21(9), 554–559. <https://doi.org/10.1073/pnas.21.9.554>

- Keselman, H. (1975). A Monte Carlo investigation of three estimates of treatment magnitude: Epsilon squared, eta squared, and omega squared. *Canadian Psychological Review*, *16*(1), 44–48. <https://doi.org/10.1037/h0081789>
- Li, J. C.-H., Nesca, M., Waisman, R., Cheng, Y., & Tze, V. M. (2021). A robust effect size for MANOVA with non-normal and non-homogenous data. *Methodological Innovations*, *14*(3), 1–12. <https://doi.org/10.1177/205979912111055949>
- Li, J. C.-H., & Tze, V. M. C. (2021). Analytic and bootstrap confidence intervals for the common-language effect size estimate. *Methodology*, *17*(1), 1–21. <https://doi.org/10.5964/meth.4495>
- McGraw, K. O., & Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin*, *111*(2), 361–365. <https://doi.org/10.1037/0033-2909.111.2.361>
- Moore, D. S., Notz, W. I., & Fligner, M. A. (2015). *The basic practice of statistics*. New York, NY, USA: Macmillan Higher Education.
- Mordkoff, J. T. (2019). A simple method for removing bias from a popular measure of standardized effect size: Adjusted partial eta squared. *Advances in Methods and Practices in Psychological Science*, *2*(3), 228–232. <https://doi.org/10.1177/2515245919855053>
- Okada, K. (2013). Is omega squared less biased? A comparison of three major effect size indices in one-way ANOVA. *Behaviormetrika*, *40*, 129–147. <https://doi.org/10.2333/bhmk.40.129>
- Pearson, K. (1923). On the correction necessary for the correlation ratio η . *Biometrika*, *14*(3-4), 412–417. <https://doi.org/10.1093/biomet/14.3-4.412>
- Ruscio, J., & Gera, B. L. (2013). Generalizations and extensions of the probability of superiority effect size estimator. *Multivariate Behavioral Research*, *48*(2), 208–219. <https://doi.org/10.1080/00273171.2012.738184>
- Vohs, K. D., Mead, N. L., & Goode, M. R. (2006). The psychological consequences of money. *Science*, *314*(5802), 1154–1156. <https://doi.org/10.1126/science.1132491>
- Wilkinson, L., & Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. [Appendix: R code]. *American Psychologist*, *54*(8), 594–604. <https://doi.org/10.1037/0003-066X.54.8.594>

Appendix

```
# data contain Group and Time
y1=data$Time[data$Group=="prime"]
y2=data$Time[data$Group=="play"]
y3=data$Time[data$Group=="control"]
y=c(y1,y2,y3)
x=as.factor(data$Group)

out=summary(aov(y~x))
ss0=out[[1]]$"Sum Sq"
df=out[[1]]$"Df"
tss0=sum(ss0)
ssb0=ss0[1]
ssw0=ss0[2]
eta2_0=ssb0/tss0
```

```
eps2_0=(ssb0-df[1]*ssw0/df[2])/tss0

bias=numeric()
B=c(100,200,500,800,1000,1500,2000,5000,8000,10000,20000)

for(R in B ){
eta2=numeric()

for( i in 1:R){
  yb=c(sample(y1,replace=TRUE),
        sample(y2,replace=TRUE),
        sample(y3,replace=TRUE))
  fit=summary(aov(yb~x))
  ss=fit[[1]]$"Sum Sq"
  eta2=c(eta2,  ss[1]/sum(ss))
}
bias=c(bias, mean(eta2) - eta2_0)
}

bias
```



Methodology is the official journal
of the European Association of
Methodology (EAM).



leibniz-psychology.org

PsychOpen GOLD is a publishing
service by Leibniz Institute for
Psychology (ZPID), Germany.