SHORT NOTES

Derivation of variance estimators and statistical inference for indices of sexual size dimorphism: an example

Szilárd Nemes* & Tibor Hartel

The Mihai Eminescu Trust, Str Cojocarilor 11 545400, Shigisoara, Romania

* Corresponding author: nemessz@gmail.com

Null-hypothesis significance testing is the primary, but not only, tool for sexual dimorphism research (1). A large variety of different indices have been developed for quantifying sexual dimorphism. Statistical tests offer better technique and a mathematically more grounded decision-making algorithm to test differences in size between the two sexes, but indices have advantages as well. Often it is easier to have a single numerical value to express the phenomenon under study. Index outcomes are interpretable as scalar expressions, and similar values always indicate similar phenomena, facilitating comparisons between studies.

One of the most frequently used sexual dimorphism indices was introduced by LOVICH & GIBBONS (2). This ratio index is rather intuitive and easy to interpret

$$SDI = \frac{\theta_x}{\theta_y} - 1$$

where θ_x represents the size parameter for the larger sex (x) and θ_y is the size parameter for the smaller sex (y). Simply put, this is size of the larger sex divided by size of the smaller sex. In order to centre the ratio around zero, one is subtracted. Usually size is quantified by sample mean or by 80th percentile. The outcome of the index is arbitrarily expressed as positive if females are larger and negative if males are larger. This index is widely used although scientists concentrate only on its calculated value without paying attention to its variability. Regardless of the method of size quantification, this SDI is a product of two random variables and therefore should itself be regarded as a random variable.

The following rather simple example will illustrate the necessity of considering the index's variance. A hypothetical study aims to quantify sexual dimorphism of a bird species that inhabits two connected forests. In the first forest 100 birds were measured, 50 females averaging 22cm, and 50 males averaging 20cm. In the second forest, due to more extreme conditions and time constraints, the researcher managed to measure only 24 birds: 20 females averaging 21cm, and only 4 males also averaging 21cm. Analyzing the collected data in the first forest the researchers observe a sexual dimorphism (SDI) of 0.1 while in the second forest the sexes exhibit no dimor-phism at all, $SDI_2=0$. As the two populations intermix, the researcher would expect the same degree of dimorphism. As the sample sizes are not equal and the sample size for the second forest is quite low, intuition leads the scientist to disregard the second forest data and only trust the index from the first. A keener look at the data set shows that the male measurements (27, 22, 19 and 14cm) from the second forest are not just too few but also have high variability and consequently lower estimation precision. This example shows the problem of incorporating not only sample sizes but also the variability in the index, and the need to consider not only the magnitude and sign of the index but also the level of confidence that the estimate has. One way to address this would be to develop Lovich & Gibbons's index to incorporate these parameters. This has serious disadvantages as it would be very difficult to incorporate sample size and variability but still preserve the intuitive simplicity and ease of interpretation. The alternative approach set forth here is to present the index with a statistic of dispersion e.g. $SDI\pm1SD$. In the following, the mathematical formula is derived for the SDI variance and the performance of the estimator is evaluated. Through a concrete example the advantages of the improved index are demonstrated.

Calculating exact variance for the Lovich & Gibbons index is difficult, although approximation techniques can be applied. An approximation commonly used is the Delta method (3), which is a procedure to find approximate mean and variance of a nonlinear combination of random variables. The delta method takes a function too complex for analytical derivation of the variance, creates a linear approximation of that function, then estimates the variance of this simpler linear function that can be used for statistical inference. The Delta method expands a function of a random variable about its mean, usually with a one-step Taylor approximation. If a function $f(x)$ has derivatives of order k, then for a constant a the Taylor series of order k about a is

$$T_n(x) = \sum_{j=0}^{k} \frac{f^{(j)}(a)}{j!}(x-a)^j$$

Statistical applications of a Taylor series are concerned with first order terms, $f(x) \approx f(\mu) + f'(\mu)(t-\mu)$. Taking expectation on both sides leads to $f(\mu)$. We can also approximate the variance of $f(x)$ by $Var[f(x_i)] \approx E[f(x_i) - f(\mu)]^2$.

Solving this equation leads to the Delta approximation for the variance

$$\sigma_\varsigma^2 = \sum_{i=1}^n \left(\frac{\partial g}{\partial x_i} \big|_{\mu_i} \right)^2 \sigma_i^2$$

A particularly useful characteristic of the Delta method is the convergence in distribution. For a given function $f(x)$ with existing first order derivative

$$\sqrt{n}[f(x) - f(\mu)] \to N(0, \sigma^2 [f'(\mu)]^2) \; ,$$

that is a normal distribution with mean zero and variance $\sigma^2 [f'(\mu)]^2$ (4).

The Delta method for the Lovich & Gibbons index leads to the following variance equation

$$\sigma_{SDI}^2 = \frac{\theta_y^2 \sigma_x^2 + \theta_x^2 \sigma_y^2}{\theta_y^4}$$

Here θ represents the size parameter, σ_x^2 the variance for the larger sex and σ_y^2 variance for the smaller sex. The variance of both sexes can be estimated from the data

$$\sigma_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Since size is quantified as the population average, the variance will be

$$\sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n}$$

The variance equation derived by the delta method is just one of all theoretically possible estimators. Most likely estimators derived by approximate methods will not be unbiased ones. Bootstrap simulations demonstrate that the bias of the derived indices is negligible (Table 1) and genuinely small. The sign of the bias has no specific pattern, thus there are no systematic errors, and no corrective applications are recommended.

TABLE 1

Bias estimation of the proposed variance index by Monte-Carlo simulation. The two samples are normally distributed with noted mean and standard deviation and different sample sizes (n and m).

Sample 1	Sample 2	Bias			
$\bar{X} + 1SD$	$\bar{X} + 1SD$	n=5, m=75	n=25, m=75	n=50, m=75	n=75, m=75
25±15	20±10	-0.0037	-0.0022	-0.0011	0.0011
25±2.5	20±1.2	0.0019	-0.0014	-0.0021	-0.0014
21±15	20±10	-0.0018	-0.0020	0.0016	0.0012
21±1.8	20±1.6	-0.0012	-0.0013	-0.0005	0.0016

Pelobates fuscus and *Bobina variageta* are two relatively widespread European toad species with pronounced sexual size dimorphism in the former and less pronounced dimorphism in the latter. Here we analyse and test their sexual dimorphism in snout-vent length using Lovich & Gibbons` SDI based on measurements from two central European populations.

Males of *B. variegata* averaged 40.81mm (±1SD:4.14) (n=44) while females averaged 40.75mm (±1SD:4.92) (n=78). The estimated dimorphism index SDI=-0.0012, while its standard deviation is 0.0209. The bias is very small 0.0001, being 26 times smaller than the estimated variance. Although it is obvious that in this case there is no sexual size dimorphism, we will use the calculated SDI and its standard deviation to demonstrate the construction of confidence intervals (CI hereafter) and null hypothesis test of the index. As a standard method we follow the normal distribution approximation

$$P\left(\widehat{SDI} - z_{\alpha/2}\sqrt{Var(\widehat{SDI})} \le SDI \le \widehat{SDI} + z_{\alpha/2}\sqrt{Var(\widehat{SDI})} \right) = 1 - \alpha$$

where z is the normal distribution quantile, while α is the desired significance level. $\widehat{SDI}$ is the estimated sexual dimorphism and SDI is the true population value. Generally $\alpha=0.05$ significance level is used, leading to z=1.96. Using the above described formula, a 95% CI from -0.039 to 0.042 is obtained. As the CI contains zero, one can be confident that the SDI for *B. variegata* does not differ significantly from zero.

A Monte Carlo simulation based on the *B. variegata* data was run to validate the build confidence interval against the generally applied computer intensive methods. A total of 1000 simulations were run. We tested the coverage probability of the build confidence interval.

Given a nominal coverage of 95% and 1000 simulations we expected that the estimated coverage level would fall in the interval 93.64 to 96.35 (5). Any value below or above the given bounds is an indication of systematic under- or over-coverage. The observed coverage probability (94.6%) of the built approximate confidence interval was slightly lower than the nominal 95% but well within the acceptance limits. As a comparison a 95% percentile bootstrap confidence interval (6) based on 1000 re-samplings resulted in a coverage probability of 94.3%. The width of the bootstrap confidence interval was slightly lower, 0.0793, than the width of the confidence interval based on our approximation, 0.0802. When comparing confidence intervals based on different methods (e.g. approximate against bootstrap intervals) we wish to choose the one with coverage probability closest to the nominal value. Between two confidence intervals, given an optimal or close to optimal coverage probability, we choose the narrowest. In our case, the proposed confidence interval and the one based on bootstrapping show no genuine difference either in coverage probability or width. Consequently, we postulate that the derived approximate confidence interval offers a valid tool as good as bootstrapping for researchers aiming to draw inference about the observed and quantified sexual dimorphism.

The probability equation for the CI easily can be transformed to a formal significance test formula

$$z = \frac{\widehat{SDI} - SDI_{H_0}}{\sqrt{Var(\widehat{SDI})}}$$

where z is the standard normal quantile and $\widehat{SDI}$ the estimated sexual dimorphism while SDI is the assumed null hypothesis value. The P-value associated with the observed z-score can be obtained from standard normal

distribution tables, or even better, from statistical software (e.g. R). In the case of *B. variegata* there is no significant difference in SDI from zero (z=0.061, P=0.950). As a comparison, a classical Student t-test for equality of means would give t=0.059 with the associated P-value of 0.953.

The measured specimens of female *P. fuscus* (n=104) averaged 63.14±4.76mm while males averaged 55.22±3.54mm (n=185). The estimated sexual dimorphism is 0.143±0.01 with a narrow 95% CI (-0.163, -0.123). The variance's bias is -0.00002, 36 times smaller than the estimated variance. The estimated SDI differs significantly from zero (z=14.28, P<0.0001). A Student t-test for equality of means also evidenced a similar result (t=16.05, P<0.0001).

It is possible to compare the difference in sexual dimorphism between the two species using the following equation

$$z = \frac{\widehat{SDI}_i - \widehat{SDI}_j}{\sqrt{\widehat{Var_i}(SDI) + \widehat{Var_j}(SDI)}}$$

The test statistic *z* has a standard normal distribution.

Comparisons made between *B. variegata* and *P. fuscus* evidenced that the magnitude of SDI is significantly larger for the latter species (z=-4.66, P<0.0001).

Since its introduction, the SDI index has consistently proved its value. With the improvements presented in this note SDI can function as a more nuanced and flexible tool. Flexibility is gained by facilitating comparison not only with a hypothetical value, but also between studies or even organisms, if needed. It also needs to be emphasized that the variability of all biological indices should be considered. The methodology used here can be applied successfully to find and assess variance estimators of biological indexes. Computer intensive methods such as bootstrapping could be employed easily both for variance estimation and statistical inference. However the Delta method offers an easy tool that can be applied to collected data directly and also data gathered from the literature. Here we chose to use the Lovich & Gibbons index, however the methodology outlined here can be used for any index of sexual size dimorphism.

ACKNOWLEDGEMENTS

REFERENCES

1. Sмiтн JR (1999). Statistics of sexual size dimorphism. Journal of Human Evolution 36:423-459.
2. Lovich JE & Gibbons JW (1992). A review of techniques for quantifying sexual size dimorphism. Growth, development, and aging, 56:269-281.
3. Oehlert GW (1992). A note on the Delta method. The American Statistician, 46:27-29.
4. Casella G & Berger RL (2002). Statistical Inference. Duxbury Advanced Learning.
5. Burton A, Altman DG, Royston P & Holder RL (2006). The design of simulation studies in medical statistics. Statistics in Medicine 25(24):4279-4292.
6. Rao TNK & Wu CFT (1988). Resampling inference with complex survey data. Journal of the American Statistical Association 83(401):231-241.