



University of Chester

This work has been submitted to ChesterRep – the University of Chester's
online research repository

<http://chesterrep.openrepository.com>

Author(s): Lewis, Stephen J

Title: A simple procedure for investigating differences in sexual dimorphism between
populations

Date: 2004

Originally published in: *InterStat*

Example citation: Lewis, S. J. (2004). A simple procedure for investigating
differences in sexual dimorphism between populations. *InterStat*

Version of item: Author's post-print

Available at: <http://hdl.handle.net/10034/56634>

A simple procedure for testing for differences in sexual dimorphism between populations

by

Stephen Lewis

s.lewis@chester.ac.uk

Department of Biological Sciences,
University College Chester,
Parkgate Road,
Chester. CH1 4BJ.
United Kingdom.

Abstract

Within an animal species, environmental factors affect males and females differently to the extent that a measure of sexual dimorphism can be regarded as indicative of the effects of environmental stresses on a population as a whole. Based upon the use of the point bi-serial correlation coefficient (an extension of the well known Pearson's product moment correlation), a combination of tests of known statistical veracity can be used to compare the sexual dimorphism of different samples. In particular, a way has been found that does not require access to raw data but permits summary statistics (means, standard deviations and sample sizes) to be used to make the appropriate calculations. This procedure has been devised with particular consideration of the needs of the statistically unsophisticated worker. In its generalized form, this procedure may also be extended to other areas of interest, such as bilateral asymmetry.

Key Words: Point Bi-Serial Correlation, Sexual Dimorphism

Introduction

Sexual dimorphism is a typical feature of most animal species. However, the degree of dimorphism expressed by a species is not constant. For traits where males usually show larger size parameters, when periods of environmental stress are encountered, sexual dimorphism decreases as males tend to fare less well than females. Alternatively, for those parameters where females tend to be larger than males, dimorphism tends to increase (Relethford and Hodges, 1985). As a result, sexual dimorphism can be used as an indicator of factors impinging upon a population as a whole.

Different statistical methods have been used to characterize sexual dimorphism. The simplest has been the expression of the mean of one sex as a percentage of the other, with ensuing tests of the statistical significance between sexes being based upon Student's *t*-test or the Mann-Whitney U-test (Hamilton, 1982). However, as Relethford and Hodges (1985) have pointed out, there is a need to test for the significance of differences in sexual dimorphism between populations as this can give an indication of differences in the environmental stresses being experienced. Few appear to have been able to achieve this to the benefit of the average worker, however. Where methods have been suggested, there have been various drawbacks – not least, the need to use computationally complex techniques requiring access to the original data. Bennett (1981) and Chakraborty and Majumder (1982), for example, considered the use of the areas under the male and female distribution curves for a given parameter, Hamilton (1982) considered the use of nested analyses of variance while Van Vark, Van Der Sman, Dijkema and Buikstra (1989) proposed two elaborate multivariate tests.

An important feature of the approach proposed by Relethford and Hodges (1985) was the requirement of basing all comparisons of sexual dimorphism on standard summary statistics (mean, standard deviation and sample size) only. In that way, previously published studies, for which the raw data are no longer available, may still be of use. Where the effects of slowly changing environmental factors on a species are concerned, it becomes possible to use data published many decades ago to make such

comparisons. However, the test they proposed, and its reformulation by Greene (1989), have been subsequently found to be simply a re-statement of the special 2x2 case of the more general test for interaction in an $r \times 2$ unbalanced ANOVA (Konigsberg, 1991).

Despite what has gone before (which has been largely confined to the field of physical anthropology), it would appear that the average biologist remains quite unaware of the potential use of a numerical expression of sexual dimorphism as an indicator of environmental factors. They also remain, in effect, without a method for none of those noted above appear to have been widely adopted. Thus, there still exists scope for considering new statistical approaches to the question of sexual dimorphism – especially where an intuitively and computationally simple approach is concerned.

An alternative approach

Instead of trying to propose a new statistical tool, an alternative approach is to use a sequence of tests of known statistical veracity using familiar terms that will have been met on most statistics courses or which may be found in any standard statistical textbook. In following this approach, the reasoning at each step will be easier for the average worker to comprehend and manage.

The data involved in work on sexual dimorphism are characterised by being mixed, consisting of continuous (ratio scale) variables associated with a dichotomous or binary (nominal scale – Stevens, 1946) attribution. Here the latter division is into 'Male' and 'Female'. A statistical test that can accommodate data with these characteristics is point bi-serial correlation.

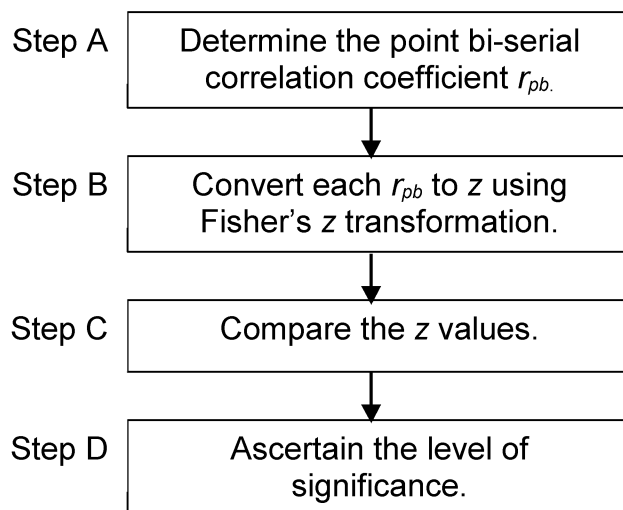
Point bi-serial correlation (Bruning and Kintz, 1987; Kendall and Stuart, 1979; Kotz and Johnson, 1982) is a lesser known extension of Pearson's product moment correlation and provides a measure of the relationship between a continuous variable, such as a size or weight parameter, and a dichotomous attribution (Bruning and Kintz, 1987; Kendall and Stuart, 1979), such as sex. As a result, the point bi-serial correlation coefficient (r_{pb}) gives, in effect, a measure of sexual dimorphism as its numerical value

represents the degree to which the continuous variable is correlated with the choice of sex as the dichotomous attribution. Two formulae for calculating r_{pb} are given in Appendix 1.

Given two such correlation coefficients, these can be compared, in the standard way, by first performing Fisher's z-transformation (Armitage, Berry and Matthews, 2001, Petrie, 1987, Bruning and Kintz, 1987) and then comparing these z values (see Appendix 2). This step of testing the equivalence of the two correlation coefficients tests whether the correlation between the chosen parameter and sex is the same in both samples. A statistically significant result, therefore, indicates that the two sexual dimorphisms differ. The procedure is quite straightforward and the steps are outlined in Figure 1.

Figure 1

For each sample:



For its simplicity and use of standard concepts and techniques, this would seem an ideal approach – particularly for the average worker, as no understanding of advanced statistical concepts is necessary and no unfamiliar ideas or complex computations have been introduced. Unfortunately, formulae for deriving the point bi-serial correlation coefficient (as given in Appendix 1) require the use of raw data to obtain the most

accurate results. In those formulae, using a pooled value for standard deviation (obtained by combining the male and female standard deviations, as given in summary statistics) can give markedly different results. Given the proviso stipulated above that, for the widest application, only summary statistics should be used, this approach would appear to be prohibited.

Another route to r_{pb}

Kendall and Stuart (1979) have shown, however, that for a given body of bi-variate data, the following relationship exists between r_{pb} and Student's t :

$$\frac{r_{pb}^2}{1 - r_{pb}^2} = \frac{(\bar{x}_1 - \bar{x}_2) \frac{n_1 n_2}{n_1 + n_2}}{\sum (x_{1i} - \bar{x}_1)^2 + \sum (x_{2i} - \bar{x}_2)^2} = \frac{t^2}{n_1 + n_2 - 2}$$

... Eq. 1

Re-arranging the two end components of Eq. 1, r_{pb} may be determined thus:

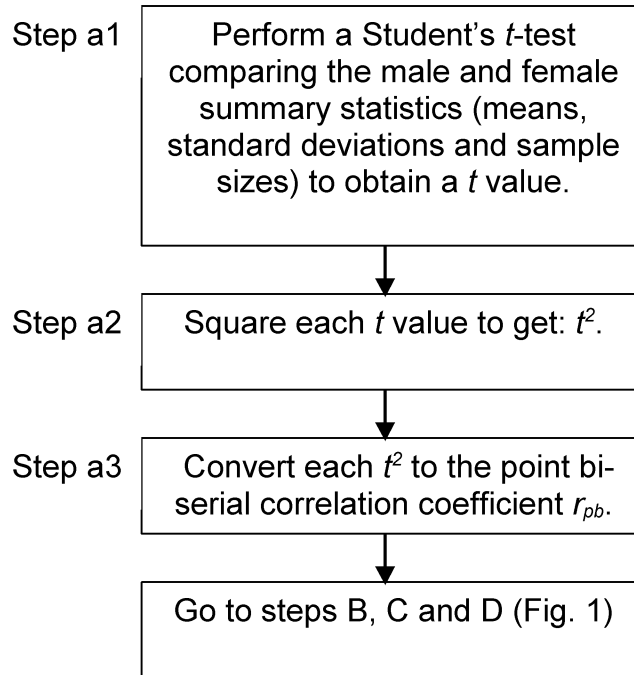
$$r_{pb} = \sqrt{\frac{t^2}{t^2 + (n_1 + n_2 - 2)}}$$

... Eq. 2

The value of t^2 can be obtained by performing a Student's t -test using male and female summary statistics and standard equations. The value of r_{pb} obtained this way does not differ from that obtainable directly from raw data were it available. Thus, where raw data do exist, the steps outlined in Figure 1 may be followed. Where summary data are to be used, additional steps prior to Step B can be followed as outlined in Figure 2. As a result, this procedure can be followed using raw data, summary statistics or a combination of both.

Figure 2

For each sample:



Extension

It should be added that the use of the point bi-serial correlation coefficient also allows the calculation of other potentially useful numerical descriptors not available via other more complex means. For example, the inverse cosine of the correlation coefficient ($\text{Cos}^{-1}(r_{pb})$) can be used to produce an angular vector (as computed in factor analysis) that can be used to give a graphical representation of the relationship between the sexes. The angular separation of two such vectors gives an impression of the difference in sexual dimorphism between the two populations investigated (Child, 1990). Also, the less commonly used coefficient of alienation ($\sqrt{1-r_{pb}^2}$) can be applied, as appropriate, to describe the *dissimilarity* between the sexes.

Conclusion

Most of the statistical techniques described here have a proven statistical pedigree and accordingly are not new in themselves. However, the use of the relationship between r_{pb} and t to obtain a measure of the correlation between two dichotomous variables from summary statistics has not been found elsewhere. The combination of steps by which this coefficient may then be used as a numerical expression of sexual dimorphism to make a comparison between samples is a novel application to this problem. Furthermore, the whole procedure is simple enough to be accessible to any worker.

As set out here, the example of sexual dimorphism has been used but this procedure is not limited to that problem alone. In its more generalized form, this procedure can be extended to other areas of interest such as the study of bilateral asymmetry, where 'Left' and 'Right' can be substituted for 'Male' and 'Female'.

Appendix 1

An equation for the point bi-serial correlation coefficient is:

$$r_{pb} = \frac{\bar{x}_1 - \bar{x}_0}{s_x} \sqrt{\frac{n_1 n_0}{N(N-1)}}$$

Where \bar{x}_1 and \bar{x}_0 are the means of the continuous variables for each category, s_x the estimated standard deviation of all the continuous variables, n_1 and n_0 are the number in each category and N the total number in both categories (n_1+n_0).

An alternative version of the point bi-serial correlation coefficient formula is:

$$r_{pb} = \frac{\bar{x}_1 - \bar{x}_0}{s_x} \sqrt{pq}$$

Where p is the proportion of individuals of one sex and $q (= 1-p)$ the proportion of individuals in the other. \bar{x}_1 , \bar{x}_0 and s_x are as before.

Appendix 2

Before comparing two correlation coefficients, each should first be transformed into a z value using:

$$z = \frac{1}{2} \ln \left\{ \frac{(1 + r_{pb})}{(1 - r_{pb})} \right\}$$

Alternatively, a shortcut, often missing from standard statistical texts, is to use: $z = \tanh^{-1}(r_{pb})$ (Porkes, 1988).

Thereafter, the two z values can be compared using:

$$\text{Test Statistic} = \frac{(z_1 - z_2)}{\sqrt{\frac{1}{(N_1 - 3)} + \frac{1}{(N_2 - 3)}}}$$

Where N_1 and N_2 are the sizes of the two samples being compared (i.e. the sizes of the male and female samples added together in each case).

This test statistic follows a normal distribution. Hence, if the result is greater than 1.96, the difference between the correlations is significant at the 5% level; if greater than 2.576, then it is significant at 1% and if greater than 3.291, then it is significant at 0.1%. (More precise values can be obtained from tables or from various computer applications such as Microsoft Excel.)

References

Armitage, P, Berry, G and Mathews, JNS (2001) *Statistical methods in medical research*. Malden, MA: Blackwell Science. (4th Edn.)

Bennett, KA (1981) On the expression of sex dimorphism. *American Journal of Physical Anthropology*. 56: 59-61.

Bruning, JL and Kintz, BL (1987) *Computational Handbook of Statistics*. Glenview (Ill.) & London: Scott, Foresman.

Chakraborty, R and Majumder, PP (1982) On Bennett's measure of sex dimorphism. *American Journal of Physical Anthropology*. 59: 295-298.

Child, D (1990) *The Essentials of Factor Analysis*. London: Cassell. (2nd Edn.)

Gray, JP and Wolfe, LB (1980) Height and sexual dimorphism of stature among human societies. *American Journal of Physical Anthropology*. 53: 441-456.

Greene, DL (1989) Comparison of t-tests for differences in sexual dimorphism between populations. *American Journal of Physical Anthropology*. 79: 121-125.

Kendall, MG and Stuart, A (1979) *The Advanced Theory of Statistics*. Vol. 2: Inference and relationship. London: Griffin. (4th Edn.)

Konigsberg, LW (1991) An historical note on the t-test for differences in sexual dimorphism between populations. *American Journal of Physical Anthropology*. 84: 93-96.

Kotz, S and Johnson, NL (1982) *Encyclopedia of Statistical Sciences* (Vol. 1). New York: John Wiley and Sons.

Petrie, A (1987) Lecture notes on medical statistics. Oxford: Blackwell Scientific Publications. (2nd Edn.)

Porkes, R (1988) Dictionary of Statistics. London: Collins.

Relethford, JH and Hodges, DC (1985) A statistical test for differences for sexual dimorphism between populations. American Journal of Physical Anthropology. 66: 55-61.

Stevens, SS (1946) On the theory of scales of measurement. Science 103: 677-680.

Van Vark GN, Van Der Sman PGM, Dijkema J and Buikstra JE (1989) Some multivariate tests for differences in sexual populations. Annals of Human Biology 16: 301-310.