Author(s): Lewis, Stephen J

Title: Further notes on a statistical method for use when investigating differences in sexual dimorphism: A discussion paper

Date: 1997


Version of item: Author's

Available at: http://hdl.handle.net/10034/67713
Further notes on a statistical method for use when investigating differences in sexual dimorphism: a discussion paper

Stephen Lewis

Abstract
A statistical method for use when investigating sexual dimorphism is described which is a development of that proposed by Lewis (1995). This development is new and remains to be fully tested. It is presented here by way of seeking constructive criticism.

Introduction
A measure of sexual dimorphism for a given population sample can be used as a single parameter to characterise the stresses affecting groups from which those samples are drawn. This has proved to be one of the most useful applications for measures of sexual dimorphism (Gray and Wolfe 1980). While van Vark et al. (1989) have proposed multivariate tests, most tests and measures of sexual dimorphism use univariate data (Bennett 1981; Chakraborty and Majumder 1982; Relethford and Hodges 1985; Greene 1989; Konigsberg 1991). These are easier to manage arithmetically and intuitively. Those tests proposed by Relethford (1985), Greene (1989) and Konigsberg (1991) have the particular advantage that they use summary statistics which may be drawn from any published source, bypassing the need to access raw data. Lewis (1997) investigated the use of point bi-serial correlation as a means of describing sexual dimorphism statistically. Point bi-serial correlation (Kendall and Stuart 1979; Kotz and Johnson 1982; Bruning and Kintz 1987) is a lesser known extension of Pearson's product moment correlation which provides a coefficient of the relationship between a continuous and a dichotomous variable, such as sex (Bruning and Kintz 1987).

The major drawback of Lewis's method was that it was necessary to use the raw data. Further investigation, however, has shown that the relationship between the point bi-serial correlation and Student's t-test may be used to overcome this limitation and provide a modified version of the previous method.

Investigation into the use of this method has, however, only just commenced and the purpose of this paper is to make the initial ideas available for comment and helpful suggestion.

Revised Method
Using summary statistics only, one wishes to derive a statistical measure of the way in which a given parameter - such as bone length - is correlated with sex, to use this to compare mixed-sex samples drawn from two different
populations.

Given access to the raw data, the point bi-serial correlation coefficient for each mixed-sex sample could be calculated directly and these coefficients compared for equivalence in the standard way. Although, in the absence of raw data, one cannot calculate the point bi-serial correlation coefficient directly, Kendall and Stuart (1979) point out that the relationship between the point bi-serial correlation coefficient ($r_{pb}$) and $t$ (from the Student's t-test) is as follows:

$$\frac{r_{pb}^2}{1-r_{pb}^2} = \frac{t^2}{n_1 + n_2 - 2}$$

Eq. 1

In the context of relating the present use, $r_{pb}$ represents the point bi-serial correlation coefficient relating the two sexes for a given population sample, $n_1$ & $n_2$ are the male and female sample sizes within the larger population sample and $t$ can be calculated from the summary statistics (means, standard deviations and sample sizes for each sex).

Given the degree of sexual dimorphism in humans, comparing males with females using a t-test for an anatomical parameter may seem unusual but here it is used simply as a means of deriving a statistical parameter (the $t$ value) from which $r_{pb}$ may be determined.

Re-arranging Eq. 1, $r_{pb}$ may be calculated as follows:

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

Eq. 2

When a value for $r_{pb}$ has been derived for each population sample, these may be compared using the equation:

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

Test Statistic

Eq. 3
where
$n_1$ & $n_2$ are the population sizes of the two mixed-sex samples being compared
and
$z_1$ & $z_2$ are each determined, given the respective point bi-serial correlation coefficients for each mixed-sex sample, using the equation:

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

Equation 4

The distribution of the test statistic follows a normal distribution, hence if the result is greater than 1.96, it 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%.

This step tests the equivalence of the two correlation coefficients and so tests whether the relationship between the two sexes for the chosen variable is the same in both samples. A statistically significant result indicates that the correlation coefficients are different and that the sexes do not show the same relationship in both samples - at least for that variable.

In common with the approach used by Relethford and Hodges (1985), the method described here is also based upon the way in which the data are correlated with sex. The statistical techniques described here are not new, although the method of deriving $r_{pb}$ from $t$ and the way in which each test is combined into a quick and simple method for producing and comparing numerical expressions of sexual dimorphism is new.

Point bi-serial correlation is a method known to relatively few non-statisticians. In its generalised form, the method described may also be applied to any system using continuous and dichotomous variables, such as the study of bilateral asymmetry.

References
Kendall, M. G. and Stuart, A. 1979 The Advanced Theory of Statistics. Griffin,


