Review

Misconceptions about logarithmic transformation and the traditional allometric method

Gary C. Packard *

Department of Biology, Colorado State University, Fort Collins, CO 80523, USA

ABSTRACT

Logarithmic transformation is often assumed to be necessary in allometry to accommodate the kind of variation that accompanies multiplicative growth by plants and animals; and the traditional approach to allometric analysis is commonly believed to have important application even when the bivariate distribution of interest is curvilinear on the logarithmic scale. Here I examine four arguments that have been tendered in support of these perceptions. All the arguments are based on misunderstandings about the traditional method for allometric analysis and/or on a lack of familiarity with newer methods of nonlinear regression. Traditional allometry actually has limited utility because it can be used only to fit a two-parameter power equation that assumes lognormal, heteroscedastic error on the original scale. In contrast, nonlinear regression can fit two- and three-parameter power equations with differing assumptions about structure for error directly to untransformed data. Nonlinear regression should be preferred to the traditional method in future allometric analyses.

© 2017 Elsevier GmbH. All rights reserved.

Contents

1. Introduction

The early part of the 20th century was marked by widespread interest among biologists in the use of simple power functions of the form

$$y = \mathbf{a}_* x^{\mathrm{b}}$$

to describe pattern in bivariate observations that follow a curvilinear path on the arithmetic (=linear) scale. The predictor variable (x) in the two-parameter equation typically was a measure of body size (e.g., body length or body mass), and the response variable (y) was some measurement taken on the structure, organ, or process of special concern. Some investigators at the time apparently fitted the equation directly to scatterplots of untransformed data by a series of trial-and-error approximations (e.g., Nomura, 1926; Kleiber, 1932), or by fitting a curve by eye and then reading from the graph (e.g., Hecht, 1913, 1916; Crozier and Hecht, 1914; Kleiber, 1932). Other workers, however, estimated the slope and intercept of a straight line drawn by hand on a graph displaying logarithmic transformations (or on a graph with logarithmic coordinates) and then took antilogs for the coefficients to obtain parameters in the power equation (e.g., Pearsall, 1927; Huxley, 1927a, b, 1932; Kunkel and Robertson, 1928). And yet a fourth group of investigators fitted straight lines to logarithmic transformations by ordinary least

squares regression and then back-transformed the resulting equation to the arithmetic scale (e.g., Clark, 1928; Galtsoff, 1931; Brody and Proctor, 1932; Green and Green, 1932). This last approach to fitting the power function continues, for all intents and purposes, to be in general use today (Warton et al., 2006; White et al., 2012; White and Kearney, 2014) and has come to be known as the traditional allometric method (e.g., Packard, 2014).

The traditional allometric method has had its critics over the years (e.g., Thompson, 1943; Smith, 1980, 1984; Lovett and Felder, 1989; Bales, 1996), and it recently has come under renewed scrutiny (e.g., Lagergren et al., 2007; Sartori and Ball, 2009; Packard, 2014, 2015, 2016, 2017a,b). Supporters of the protocol are understandably concerned that a large body of published research might be undermined if criticisms of the method were taken seriously, and they consequently have mounted a spirited defense of their research paradigm (e.g., Klingenberg, 1998; Nevill et al., 2005; Kerkhoff and Enquist, 2009; Xiao et al., 2010; White et al., 2012; Ballantyne, 2013; Glazier, 2013; Lai et al., 2013; Mascaro et al., 2014; Niklas and Hammond, 2014; Lema&tre et al., 2015). However, the defense is based in many instances on ill-defined arguments and/or misunderstanding of various statistical methods. Here I examine four of the most common misconceptions.

2. Misconceptions about bivariate allometry

2.1. Form of the allometric equation

Philip Gingerich has proposed that "the allometric equation is not a power function of x and y as is so often stated, but rather a linear function of log x and log y" (Gingerich, 2000, p. 220). The suggestion was based on his study of normal vs. lognormal distributions for random error (i.e., residuals) in samples of biological data and not on an explicit analysis of allometric variation. The assertion consequently rests on something of an extrapolation, but it may, nonetheless, describe contemporary research on allometric variation fairly accurately. Logarithmic transformation was used in the 1920s and 1930s to fit simple, two-parameter power functions to observations expressed on the arithmetic scale. As time went on, however, transformation became less and less a means to fit a power function to untransformed data, and the equation of simple allometry simultaneously became more and more a justification for performing the transformation. Thus, contemporary analyses of allometric variation typically begin with a token nod to the equation of simple allometry, proceed with the rote transformation of data to logarithms, and then continue with the fitting of a straight line to the new distribution (Smith, 1980, 1984). Validation of the fitted model usually is limited to a graphical display of the equation in log domain, and R^2 may also be reported as a measure of goodness of fit. However, the quality of the model seldom is assessed graphically on the original scale (Packard, 2017b), and R^2 for the fit to logarithms has no bearing on strength of the relationship between the untransformed variables. Thus, it is easy to believe that the allometric equation in current practice is, in fact,

$$\log(y) = \log(a) + b_* \log(x).$$

If the allometric equation actually is a linear function of $\log(y)$ and $\log(x)$, future reports of allometric variation should be framed differently. For instance, reports of new research need not (and probably should not) begin with a reference to a two-parameter power function, because the power function is neither necessary nor relevant to the analysis; and no attempt needs to be made at interpretation on the original scale, because the original scale also is irrelevant. These recommendations are consistent with the common belief that observations in logarithmic form are at least as meaningful as untransformed values (e.g., Peters, 1983; Kerkhoff

and Enquist, 2009; Glazier, 2013; Lai et al., 2013), and the relationship of putative interest is, after all, between $\log(y)$ and $\log(x)$.

It is unclear how this relationship between $\log(y)$ and $\log(x)$ is to be interpreted in biologically meaningful terms when the findings cannot be placed in the context of the original measurements (e.g., Reyment, 1971; Finney, 1989; Osborne, 2002; Feng et al., 2013). This problem of interpretation is an unavoidable consequence of accepting at face value the aforementioned premise about the allometric equation. Gingerich had an important point to make about the kind of random error that may occur in biological data, but his characterization of the allometric equation should not be taken literally. The equation of simple allometry is

$$y = a_* x^b,$$

as so often stated (e.g., Huxley and Teissier, 1936; Huxley, 1950), and the linearized expression

$$\log(y) = \log(a) + b_* \log(x)$$

merely provides a way to estimate parameters in the allometric equation via an intermediate step involving logarithms (e.g., West and West, 2012). In other words, a straight line fitted to logarithms is a means to an end and not an end in itself. The linearized expression was an essential tool early in the last century, because neither statistical theory nor statistical practice was sufficiently advanced at the time to permit fitting the power function directly to untransformed observations. However, the equation now can be fitted to untransformed data by nonlinear regression, and the issue of normal vs. lognormal error can be addressed simultaneously by the computational algorithm (Packard, 2015, 2016, 2017a).

2.2. Proportional change and relative growth

Several investigators – all of whom cite Huxley (1932) for their rationale – have argued that logarithmic transformation (i.e., the traditional allometric method) is necessary in allometric research because allometry is all about proportional change and relative growth (e.g., Kerkhoff and Enquist, 2009; Glazier, 2013). But what, exactly, is meant by "proportional change" and "relative growth," and is logarithmic transformation really necessary for describing these processes?

Julian Huxley's treatise on "Problems of Relative Growth" (Huxley, 1932) focused on how some part of the body changes in its proportion to the body as a whole as both the part and body increase in size. At the time of Huxley's writing, proportional (or relative) size of a part was commonly expressed as a percentage of size of the body, and relative growth was reflected in the changing percentage as the animals (or plants) increased in overall size (e.g., Huxley, 1924a). This expression of relative growth could be represented graphically (see Fig. 2 in Huxley, 1932), but the actual relationship between the structure of interest and body size could not be quantified accurately owing to x being a component of both the predictor and response variables (see Pearson, 1897; Atchley et al., 1976; Albrecht et al., 1993; Kronmal, 1993).

Huxley's desire to express the relationship between x and y mathematically is what led to his independent discovery of the formula for simple allometry (Reeve and Huxley, 1945), which provides a simple, yet explicit, mathematical description for the relationship between the variable of interest and body size. The two-parameter power function can be rearranged algebraically to show that the response variable, y, varies as a constant proportion of x^b (e.g., Newcombe, 1948; White et al., 2012), or

$$a = y/x^b$$

where a and b are fitted constants. Because y maintains a fixed proportional relationship to x^b as both y and x increase in size,