

Frog morphometrics: a cautionary tale

Lee-Ann C. HAYEK*, W. Ronald HEYER**¹ & Claude GASCON***

* Mathematics & Statistics, MRC 136, National Museum of Natural History,
Smithsonian Institution, Washington, DC 20560-0136, USA
[hayek.lee-ann@nmnh.si.edu]

** Amphibians & Reptiles, MRC 162, National Museum of Natural History,
Smithsonian Institution, Washington, DC 20560-0162, USA
[heyer.ron@nmnh.si.edu]

*** Field Support Program, Conservation International,
2501 M Street, NW, Suite 200, Washington, DC 20037, USA
[c.gascon@conservation.org]

Scant attention has been paid to measurement error in frog morphometric studies. We study both interobserver effects of measurement on the same specimens of *Vanzolinius discodactylus* (Anura, Leptodactylidae) and intraobserver effect of repeated measurements on a single *V. discodactylus* specimen. Interobserver measurements differ statistically and result in different biological interpretations in some cases. Evidence is provided that log transformation of raw data is often unnecessary. Allometric transformation of measurement variables to remove size effect requires parallel regression slopes of variable against size. This requirement is not met with the *V. discodactylus* data, nor is it likely to be met when several variables are used in a morphometric study. We recommend: assume measurement differences between sexes in frogs and analyze data separately by sex; consider and select the most appropriate statistical model options for data analyses; avoid pseudoprecise measurements; do not rush to logarithmic transformation; remeasure at least one individual frog 20 times to provide an assessment of measurement error in data interpretation; be conservative in drawing biological inferences from morphometric analyses, basing interpretations and conclusions only on very robust effect size estimates and differences.

INTRODUCTION

Frogs are relatively soft-bodied organisms and their preservation requires considerable care. Limbs and body must be correctly positioned to achieve standardized preparation. Unfortunately, different preservatives and different individual techniques result in very different museum preparations for the same species (fig. 1). Therefore, precise, comparable measurements of preserved frogs are difficult. For example, one of the standard measurements taken on frogs, snout-vent length (SVL), is somewhat problematic in larger preserved frogs, because the sacral-urostyle portion of the body usually is fixed at an obtuse angle to the vertebral column. How much one "straightens out" the preserved animal has an effect on the

1. Corresponding author.

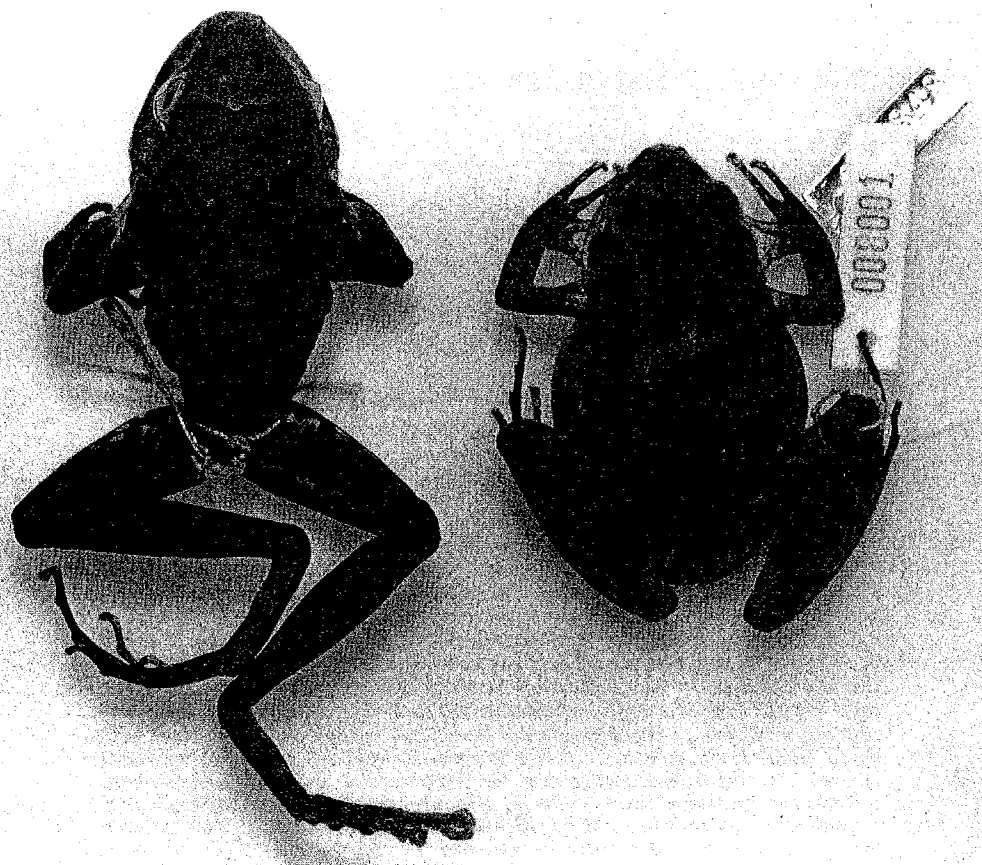


Fig. 1. - *Thoropa miliaris* (USNM 38936 on left, USNM 229848 on right) showing preservation/positioning differences that make accurate, comparable measurements difficult.

resultant measurement. In spite of (or, perhaps oblivious to) these difficulties, researchers have used frog measurement data to address a variety of scientific questions. There has been little attention paid to precision and repeatability of frog measurement data and how this variation might affect the scientific questions being addressed.

We know of only one study (LEE, 1982) that demonstrated important measurement differences between fresh and preserved frogs and differences in measurements taken on the same individuals at the same state of preservation. In that study, Lee took all the measurements himself using the same measuring equipment and methodology throughout. Although LEE (1982) presented extensive literature on the effects of preservation technique on fish morphology and discussed its relevance to frog morphometrics, herpetologists have generally ignored his warnings.

We are not aware of any published studies of the effect of different individual researchers taking the same set of measurements on the same frogs to measure inter-observer variability (although A. Dubois and A. Ohler have unpublished data on this topic, personal communication). Studies on other groups of organisms demonstrate that such differences are not trivial. LEE (1990) found differences in precision between two observers on scale count data taken from the same lizards. YEZERINAC et al. (1992) found that measurement error varied considerably, depending on the variable, for bird skeleton measurement data. In these studies, a constant value was being measured. That is, the number of scales did not change on any individual lizard, nor did the individual bird bones change size or shape. As indicated above, this is not true for whole frog specimens: how the specimen is positioned will determine what the value of the measurement will be for several of the measurements (variables) commonly taken for frog morphometric studies.

PAGANO & JOLY (1999) compared a select group of morphological measures on water frogs with an analysis of allozymic markers. These authors concluded that frog morphology was of limited use for their identification purposes. They determined frog body landmarks for measurement points from digitized photographs of specimens. Data were input and analyzed on a computer. Similar methodology has proved acceptable for characterization of stratigraphic sections (see e.g., BENSON et al., 1995), in which the surfaces are approximately linear and two-dimensional. However, for examination of three-dimensional, soft-bodied organisms, the use of such methods further complicates the measurement process. Despite the stated advantage of magnification of digitized figures for measurement purposes, statistical error minimization has not been proved to be achievable for measurements taken from frog photographs. Based on our experience, we do not recommend using photographs of frogs from which to take morphometric data.

One of us (CG) took a series of measurements on specimens of the frog species *Vanzolinius discodactylus* (Anura, Leptodactylidae) from the Rio Juruá in Brazil to test the riverine hypothesis of speciation (GASCON et al., 1996). Another of us (WRH) used the same specimens in a study examining differentiation throughout the entire species range of *V. discodactylus* (HEYER, 1997). WRH took the same set of measurements on the same frogs that CG measured. The two data sets were given to LCH to analyze and evaluate. During the course of this study, LCH reevaluated the statistical procedures and assumptions used in the GASCON et al. (1996) study.

The objectives of this study are: (1) to evaluate inter- and intra-observer statistical differences of measurement sets; (2) to understand the kinds of differences investigators create when measuring frogs; (3) to evaluate the effect of measurement differences on certain statistical procedures that are generally applied in frog morphometric studies; and (4) to judge whether measurement differences yield different biological interpretations.

METHODS AND MATERIALS

Fourteen measurements were made on each frog, following the methodology in GASCON et al. (1996). The fourteen variables are: snout-vent length (SVL), nostril separation, eye

width anterior, eye width posterior, head width, head length, eye to nostril distance, tympanum diameter (tympanum height of GASCON et al., 1996), eye length, thigh length (femur length of GASCON et al., 1996), shank length (tibia length of GASCON et al., 1996), foot length, maximum width of disk on third finger, and maximum width of disk on fourth toe.

Prior to WRH's taking of these data, he confirmed landmarks with CG for a subset of the variables in an attempt to make certain that the measurements would be comparable.

CG and WRH measured each individual one time.

CG used digital calipers linked to an IBM-PC; measurements were made to the closest 0.01 mm and the data were recorded with three decimal places. WRH used Helios dial calipers; measurements were made to the closest 0.1 mm and the data were recorded with one decimal place.

To assess individual measurement error, WRH measured one male, USNM 348976, 20 times over a 12 day period. The eye region on one side of the head is slightly squashed, otherwise this specimen is in reasonable shape. The specimen is about average in overall state of preservation and positioning in terms of ease of measurements. Measurements were taken at various times of the day and measurements were never taken one immediately after the other to eliminate or minimize carry-over effects of learning or memory. For SVL, efforts were made to focus visually on the caliper jaws when measuring the specimen and not to look at the readout dial until after the jaws had been set. All other measurements were taken under a dissecting microscope with the calipers while the measurement readout dial was not visible in the field of observation. Measurements were recorded on dated and timed separate, individual data sheets.

CG and WRH used different criteria to categorize sex of the individuals. CG used three categories: F, M and 0. In cases where CG opened the frog to take tissues, sex and whether the individual was adult or not were determined by the state of its gonads. Individuals recorded as 0 were not opened. These data were recorded under field conditions. For the morphological analyses reported by GASCON et al. (1996), data for adult and non-adult males were combined as were the data for adult and non-adult females. WRH used five categories: M, F, B, G and J. The M (adult male) category was determined by presence of vocal slits in males. The F (adult female) category was determined by presence of developed ova or some curliness of the oviduct in females. The B (juvenile male) category was determined by presence of testes. The G (juvenile female) category was determined by presence of ovaries. The J (juvenile) category was used when sex could not be determined, either because the gonads were indeterminate in very small specimens or the gonads had been removed from the specimens when tissues had been taken. These data were taken in the laboratory with the aid of a Wild stereoscopic dissecting microscope.

Male and female immature gonads of *Vanzolinius discodactylus* are quite similar in appearance and difficult to differentiate without detailed examination under magnification. Both ovaries and testes have a mosaic-like pattern externally. The only consistent difference between immature gonads is that the testes have a smooth external surface, whereas ovaries have an irregular external surface. Not surprisingly, the difficulty of differentiating gonads using the unaided eye resulted in several different interpretations of sex by CG and WRH. The differences are (CG determination, followed by WRH determination): INPA 2410 (F, B); INPA 2371, 2433, 3397, 5605, 5671, 5728, 5735, 5736, 5799, 5801 (M, G); INPA 3572, 5571 (F,

J, gonads now removed in both); INPA 3177, 3573, 5524, 5592, 5670, 5697, 5730 (M, J, gonads now removed in all).

WRH's categories of adult male (M) and adult female (F) are used in the analysis section for both the CG and WRH measurement data sets unless otherwise noted. Using this categorization, 88 adult individuals are available for analysis. Each variable was examined and summarized separately for male and female adults. Graphs and descriptive statistics were calculated and assumptions tested prior to means tests or predictive analyses. Logarithmic transformations were performed and descriptive statistics calculated on the transformed values as well. Tests of normality were performed and discussed below.

In this study, we cannot calculate residual measurement error because we do not have the "true" value of the variable for any individual specimen. Similarly, we are unable to assess a statistical variability estimate for the factors involved in the overall measuring error. That is, we cannot remove intra-observer variability from inter-observer measurement error. We therefore evaluate the two factors separately.

We distinguish "precision" from "accuracy". Accuracy is the closeness of an observer's measurement to the quantity intended to be measured. In our case, this is unknown for the true value of the frog's morphological measurement but can be evaluated by considering the closeness of the results of the two observer's values. Precision refers to the entire class of measurements and how well repeated measurements self-conform. In this case, the mean value does not have to be the "true" value of the variable. To examine these characteristics we calculated both inter- and intra-observer variability estimates and also descriptive measures for qualitative evaluation of the frog data.

Data were analysed either using direct mathematical formulae or using the software package SPSS 8.0 (ANONYMOUS, 1998). Although the discriminant function analyses were done using SPSS 8.0 (ANONYMOUS, 1998), the figures were produced using either SYSTAT versions 7 (ANONYMOUS, 1997, for fig. 7) or 9 (ANONYMOUS, 1999, for fig. 5-6).

THE APPROPRIATENESS OF RAW DATA TRANSFORMATION PROCEDURES IN FROG MORPHOMETRIC STUDIES

GASCON et al. (1996) used an allometric transformation procedure described by THORPE (1976) in an effort to remove size effects from the data. The Thorpe procedure (presented in detail in THORPE, 1975) involves two steps: (1) log-transforming the original measurement data; and (2) transforming the log values using a common slope based on the entire data set. The topic of transforming raw data is discussed first, followed by demonstration that the statistical assumptions of the Thorpe procedure are not met by the *Vanzolinius* data as used by GASCON et al. (1996).

Although not specifically mentioned by GASCON et al. (1996), the raw measurement data were log-transformed as part of THORPE's (1976) transformation procedure. Raw data are transformed as a matter of course in many multivariate analyses of frog morphometric data (for a recent example see GREEN et al., 1997). SOKAL & ROHLF (1969) state that log transformation is the most common transformation for biological data and they provide a cogent

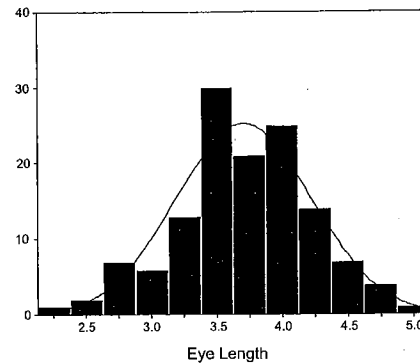


Fig. 2. – Histogram of eye length values measured by CG on total sample of 131 frogs with normal distribution best fit.

discussion on the topic of log-transforming variables as a way to meet some statistical test assumptions that are not met by raw variable data. However, this transformation is often applied routinely, when, in fact, it may be either unnecessary or incorrect to do so.

Replacing each measurement by its logarithm may result in more approximate variance equality. Also, for many biological applications the data can be normalized by this change. The assumption of concern for our purposes is whether the variables are normally distributed. Using BESTFIT (ANONYMOUS, 1995) on the data as analyzed by GASCON et al. (1996), untransformed variables for the entire sample size of 131 individuals were fit with a normal distribution (see fig. 2 for an example). We used the Anderson-Darling test criterion as well as a chi-square test of fit. The Anderson-Darling criterion is more tail-sensitive than the ordinary chi-square goodness-of-fit test.

SOKAL & ROHLF (1969) state that the log transformation may be appropriate and useful when the means of the samples are proportional to the range or standard deviation of the respective samples. The biological questions we are asking of the *Vanzolinius* data require grouping of the data by locality. None of the variables, for the total sample or when organized by locality, show a relationship of mean with either standard deviation ($r = 0.06$ ns) or range ($r = 0.19$ ns). In addition, each raw variable plot shows approximate symmetry, lack of prominent skewness and unimodality (for example, snout-vent length as shown in fig. 3).

Thus, the data as analyzed by GASCON et al. (1996) can be appropriately analyzed as raw variable measurements, rather than log-transformed variables. It is not incorrect statistically to apply and use the logarithmic sample data for this problem. It is, however, unnecessary for the morphological variables being measured here.

The reason GASCON et al. (1996) used logarithmic transformation was to attack the problem of allometry effects in their data, which included both adults and juveniles. THORPE (1976) presented a procedure that uses a log transformation as an initial step toward eliminating the influence of allometry. We examined the application of this approach and found it inappropriate for the *Vanzolinius* data for the following reason.