Global Journal of Science Frontier Research: I
Interdisciplinary
Volume 22 Issue 1 Version 1.0 Year 2022
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 2249-4626 & Print ISSN: 0975-5896

A Numerical Criterion for Assessing the Discriminative or Aggregative Potential of a Taxonomic Character. Part II

By Joaquim Carlos Sena Maia
Universidade Federal do Paraná

Summary- An alternative method to assess the aggregative/discriminative potential of taxonomic characters for continuous variables obtained through measurements.

Keywords: numerical method: taxonomy; potential, aggregative, discriminative; character.

GJSFR-I Classification: DDC Code: 571.876 LCC Code: QL639.25

A Numerical Criterion for Assessing the Discriminative or Aggregative Potential of a Taxonomic Character. Part II

Joaquim Carlos Sena Maia

Summary- An alternative method to assess the aggregative/discriminative potential of taxonomic characters for continuous variables obtained through measurements.

Keywords: numerical method: taxonomy; potential, aggregative, discriminative; character.

I. Introduction

The evaluation of the aggregative or discriminative potential of a biological characteristic (anatomical, morphological, behavioral, etc.) allows estimating the contribution of this characteristic in the formation and systematic delimitation of the taxonomic units where it is used. Maia (2021) presented a method for making this assessment, estimating the potential on a scalefrom 0 to 5 and working with discrete variables obtained through attribute counting and morphological variations coding. Similarly, in the present work, we propose a methodology for continuous variables obtained through measurements.

II. Material and Methods

The material used refers to measurements performed on 6 species (n = 6) of the genus Plebeia that occur in southern Brazil: *P. juliani, P. meridionalis, P. droryana, P. emerina, P. remota* and *P. saiqui,* represented in the tables by the letters J, M, D, E, R, S (Maia, 2017). To exemplify the methodology proposed here, we will use measurements (mm) of the width of the jaws of 5 specimens (m = 5) of each species, as shown in Table 1.

Table 1: Width of the jaw of six species of the genus Plebeia

J	R	S	D	E	M
0.90	1.40	1.20	1.15	1.50	1.00
1.00	1.80	1.30	1.20	1.30	0.90
1.00	1.40	1.50	1.15	1.20	0.90
0.90	1.45	1.30	1.20	1.20	0.70
0.90	1.40	1.30	1.00	1.20	0.80

Initially, the data are subjected to one-way analysis of variance (ANOVA) and we interpret the test value (F=20.50 - P<0.05) noting that there are significant differences among the analyzed samples, as shown in Table 2

Table 2: ANOVA of Width of the jaw of six species of the genus Plebeia

Source of variation	SS	DF	QM	F
Total	1.763	29		
Among groups	1.434	5	0.287	20.50
Within groups	0.324	24	0.014	

In the next step we will calculate the minimum significant difference (msd) using the standard deviation "within groups", $S_E = \sqrt{0.014}$ (also called standard error), and a tabular value (q) of the Tukey Test, in order to assess the significance of the differences between the sample means.

$$\text{Calculation of msd in this example: msd} = q.\,(S_E / \sqrt{m})$$

$$\text{msd} = 4.37\,(0.118 / \sqrt{5}) = 0.23$$

Table 3: Means of jaw widths and msd

M	J	D	E	S	R
0.86	0.94	1.14	1.28	1.32	1.49
msd = 0.23					

Author: Professor Adjunto do Departamento de Estatística da Universidade Federal do Paraná – Brazil. e-mail: jcsenamaia@gmail.com

The total number of pairs (JM, JD, RS...) is obtained by calculating the combination $\binom{n}{2}$.

In the case of 6 species, we will have 6(6-1)/2 = 15 pairs corresponding to the cells in Table 4.

For example, for the pair JM, we calculate the difference between the means $\overline{X}j$ and $\overline{X}m$ and compare it with the msd. If the difference is less than the msd, we accept that the jaw width is statistically equal for both species. The pair is coded as 0 (zero).

The 15 possible comparisons are shown in Table 4 and the zeros in the crossings of the lines and columns corresponding to the letters (M,J,...,R) show that for those pairs, the jaw width character does not present significant differences. The number of zeros in table 4, that is, the number of pairs whose means do not differ significantly, is represented by the letter Z. In this example

Table 4: Matrix of similar pairs. (0=non-significant difference)

	M	J	D	E	S	R
M						
J	0					
D		0				
E			0			
S			0	0		
R				0	0	

Thus, we can calculate the degree of discrimination (D) of the jaw width character through the formula:

$$D = 5(1 - 2Z/n(n-1))$$

In this example we would have:

$$D = 5\left(1 - \frac{2(7)}{6(6-1)}\right)$$

$$D = 2.67$$

The interpretation of the D value can be done as follows:

$0<=D<=1$-very aggregative characters
$1<D<=2$-aggregative characters
$2<D<= 3$-intermediate characters
$3<D<= 4$-discriminative characters
$4<D<= 5$-very discriminative characters

III. RESULTS AND DISCUSSIONS

A traditional taxonomic study, which includes both the identification of species and their grouping, forming taxonomic units of a higher hierarchical level (genera, families, etc.), uses the analysis of characters (morphological, anatomical, behavioral, etc.) that guides this ordering.

Most characters contribute to both aggregate and discriminate, in proportions that vary for each character, as demonstrated by Maia (2021)

Taxonomy has been modernized since the creation of Numerical Taxonomy in the 1950s with the pioneering work of Sneath (1957 and 1958) and Michener and Sokal (1957). After that, with the development of new numerical analysis methods such as Principal Component Analysis, Cluster Analysis and others, taxonomists assimilated these new tools.

Ranjana et all (2013) found that traditional taxonomy based on morphology has often failed to accurately identify species due to the occurrence of cryptic species that are reproductively isolated but morphologically identical. They used Discriminant Function Analysis (DFA) and Cluster Analysis and the results show that DFA is a reliable statistical tool for identification and that Cluster Analysis works effectively for classification and species identification.

More recently, the use of DNA to identify and classify species has been widely used and has shown great superiority over traditional methods.

Pfenninger et all (2006) compared the suitability and effectiveness of traditionally used shell morphology with DNA-based methods to distinguish between species of the freshwater snail genus Radix (Basommatophora, Pulmonata).

They concluded that the taxonomic distinction of species of the genus Radix should not be based on the morphology of the shells, as their variability is: i) continuous, ii) largely overlapping between different species and iii) phenotypically plastic in response to environmental conditions. They stated that the greatest conceptual advantage of DNA taxonomy over morphological methods lies in the direct inheritance of the characters used for identification. Unlike many morphological characters, DNA sequences are not subject to potentially misleading environmental or developmental modifications.

From a taxonomic point of view, a given biological character may have an aggregative or a discriminative potential, which can be quantitatively evaluated.

We believe that a preliminary assessment of the possible taxonomic contributions of each character can help in choosing the variables that will be later used in more sophisticated and costly analyses.

Maia (2021) developed a method to quantify this potential in discrete variables (attribute counts, coding of morphological variations, etc) on a scale between zero and 5.

In the present work we deal with the evaluation of this potential in the continuous variables obtained through measurements.

The discrete variable has contours clearly established and can be easily interpreted by counting attributes or by coding qualitative properties (colors, shapes, etc). In continuous variables, the limits of possible groups need to be statistically calculated (Maia, 2017).

The analysis of the differences between the arithmetic means of the studied groups and the construction of confidence intervals allow identifying the formation of possible groups of species whose means do not differ significantly. In this work, we used Tukey's test, calculating the minimum significant difference msd = 0.23 (Sokal&Rohlf,1981). When interpreting the ANOVA results, it is common to order the means in ascending order and underline those that do not differ significantly. (Gomes, 1970)

$$\begin{array}{cccccc} \mathrm{M} & \mathrm{J} & \mathrm{D} & \mathrm{E} & \mathrm{S} & \mathrm{R} \\ 0.86 & 0.94 & 1.14 & 1.28 & 1.32 & 1.49 \\ \cline{1-2} \cline{3-5} & & & & & \\ \cline{2-3} \cline{4-6} \end{array}$$

These underlined means are statistically equals and coded with zero in Table 4. In this example, of the 15 pairs of possible comparisons, 7 can be considered similar. The discriminative potential of the jaw width characteristic is represented by the 8 comparisons in which the differences between the means are greater than msd

Direct application of the Tukey test is recommended when the samples have the same size, as in our example, m = 5. If one or more samples do not meet this specification, the Tukey-Kramer test must be used to calculate the msd. It should be remembered that Analysis of Variance should only be used after evaluating the requirements for independence, normality and homogeneity of variances

In order to allow comparisons between similar analyses, the results can be standardized on a scale between zero and five, whose result D is obtained by calculating the variation of an index that uses the minimum and maximum values within the analyzed group. The aggregative potential (D') is calculated complementary, D'= 5 - D, since aggregation and discrimination are complementary concepts. In this example, D=2.67 suggests that the jaw width character belongs to the intermediate type as it both aggregates and discriminates

Table 5: Degree of Discrimination

0 <= D <= 1 - Very little discriminative character
1< D <= 2 - Little discriminative character
2 < D <= 3 - Moderate discriminative character
3 < D <= 4 - Discriminative character
4< D <= 5 - Very discriminative character

Table 6: Degree of aggregation

0 <= D' <= 1 - Very little aggregative character
1< D' <= 2 - Little aggregative character
2 < D' <= 3 - Moderate aggregative character
3 < D' <= 4 - Aggregative character
4< D' <= 5 - Very aggregative character

IV. CHARACTER WEIGHTING

The values of D and D' can be used as a weighting criterion (for both discrete and continuous variables) when we want to assign different values to some biological characteristics.

The use of weights to rank character contributions in a taxonomic work is quite controversial. There are those who defend it and those who criticize it, since the times when Numerical Taxonomy was established as an alternative method for zoological and botanical classification, such as:

- Sneath (1973) states that the main conceptual difficulty that retarded the progress of the Numerical Taxonomy was the problem of the weighting of characters.
- Michener and Sokal (1957) conclude that, even if it is desirable, there is no rational way to determine character weights and, in practice, we should assign the same weight to all.
- Burtt(1964) warned that numerical taxonomy need not necessarily assign equal weights to all characters (isocratic classification).
- Farris (1966) suggests that characters that vary little within populations are more reliable indicators of cladistic relationships than characters that vary more, and therefore should have greater weights.
- Goodman (1969) believes that a character should be inversely weighted to the variance within the taxonomic unit.
- We do not intend to go into the merits of the disagreements. We just point out that the value D (and D') can eventually be used as a weighting criterion by those who defend this practice.
- Maia (2021) created tables for the values of D and D', but due to a graphical composition error, the tables only showed the amplitudes of variation of the weights and not the respective limits. This flaw is

now fixed and so we present the two tables as they should have been published.

V. CONCLUSIONS

- A methodology is presented to assess the discriminative or aggregative potential of continuous variables.
- Analysis of variance (one-way) makes it possible to identify whether there are significant differences among the samples tested and the contribution of the analyzed character to grouping or separating the species.
- The formation of groups whose species are statistically equals for a given characteristic is done through the Tukey Test applied to the means of the measurements of the samples.
- The discriminative/aggregative potential of continuous variables can be expressed on a scale that varies between zero and 5, just as to discrete variables.
- The values of D and D' can be used as weighting criteria to rank the contributions of characters in taxonomic studies.

ACKNOWLEDGMENTS

To Dr. Sebastião Laroca for suggestions, to Dr. Gabriel Maia Braga for the text translations and to Dr. Terezinha Iza Cezar for the final review and corrections.

REFERENCES RÉFÉRENCES REFERENCIAS

1. BURTT, B. L. 1964. Angiosperm taxonomy in practise. *Phenetic and Phylogenetic Classification,* pp 5-16. Sys. Ass. Pub. 6. 164 pp in Sneath P. H. A. & R. R. Sokal. 1973. *Numerical Taxonomy.*
2. FARRIS J. S. 1966. Estimation of conservatism of character by constancy with biological populations. *Evolution 20: 587-591.*
3. GOMES, F.P. 1970. Curso de Estatística Experimental. *Universidade de São Paulo. Escola Superior de Agricultura Luiz de Queiroz. 429 pp.*
4. GOODMAN, M. M. 1969. Measuring evolutionary divergence. *Jap.J.Genet. 44 (suppl. 1): 310-316 in Numerical Taxonomy.*
5. JAISWARA, R., Nandi D., Balakrishnan R. 2013. Examining the effectiveness of discriminant function analysis and cluster analysis in species identification of male field crickets based on their calling songs. *PLoS One. 2013 Sep 25;8(9):e75930. doi: 10.1371/journal.pone.0075930.eCollection 2013*
6. MAIA, J. C. S. 2017. Análise estatística interespecífica das espécies paranaenses (Brasil) de Plebeia (Apidae, Meliponini). *Acta Biol.Par., Curitiba, 46(1-2): 39-58.*
7. MAIA, J.C.S. 2021. A numerical criterion for assessing the discriminatory/aggregative potential of a taxonomic character with a Fortran 90 Program for calculations. *Acta Biol.Par.Curitiba, 50 (1-4):7-17.2021.*
8. MICHENER, C. D. & R. R. Sokal. 1957. A quantative approach to aproblem in classification. *Evolution 11:130-162.*
9. PFENNINGER M., Cordellier, M. & Streit, B. 2006. Comparing the efficacy of morphologic and DNA-based taxonomy in the freshwater gastropod genus Radix (Basommatophora, Pulmonata). *BMC Evol Biol 6, 100.*
10. SNEATH, P. H.A. 1957. The application of computers to taxonomy. *J. Gen. MIcrobiol.,* 17, 201-226.
11. SNEATH, P.H.A. 1958. Some aspects of Adansonian classification and of the theory of correlated features. *Ann. Microbiol. Enzimol., 8, 261-268.*
12. SNEATH, P. H. A. & R. R. SOKAL. 1973. *Numerical Taxonomy, W. H. Freeman and Company, San Francisco 573 pp.*
13. SOKAL, R.R. & F.J. ROHLF. 1981. *Biometry, W.H. Freeman and Company, New York 429 pp.*