MedCrave Step into the World of Research

MedCrave Online Journal of Biology and Medicine

Research Article

An objective criterion for weighting taxonomic characters

Abstract

In this work we present the programs developed to calculate, through a weighting system, the contributions of taxonomic characters to aggregate species in higher taxa or to individualize them in identification processes. The concepts established in previous works are presented graphically in order to facilitate the understanding of the concepts of aggregative and discriminative potentials of a character. A step-by-step tutorial is presented to facilitate the use of the programs developed by the author to calculate the weighting of each character, numerically translating the ability to aggregate and discriminate taxonomic units, by means of statistical analysis.

Keywords: weighting, aggregating, taxonomic characters, potential

Volume 9 Issue 2 - 2024

Joaquim Carlos Sena Maia
Adjunct Professor (retired) of the Department of Statistics at UFPR, Brazil

Correspondence: Joaquim Carlos Sena Maia, Adjunct Professor (retired) of the Department of Statistics at UFPR, Brazil, Email jcsenamai@gmail.com

Received: April 12, 2024 | **Published:** May 22, 2024

Introduction

The use of weights to emphasize the greater or lesser taxonomic importance of a biological character was very prominent with the emergence of Numerical Taxonomy in the 1950s. If, on the one hand, the use of weights is very appropriate for numerical methods, on the other hand, whatever method is employed, it is important to know the value of the participation of each character in the formation of the taxons studied. As the determination of weights involves calculations, including statistical analysis, we present in this work a program developed in FORTRAN 90 that contemplates all the routines of the necessary calculations.

Material and methods

We will use the same material published by Maia et al.[1] used later in the publications that described the methodologies of the calculations of the weights for discrete variables (counts, coded attributes, etc.), Maia et al.[2] and for continuous variables (measurements), Maia et al.[3] in order to facilitate the joint analysis of these three works that complement each other in the study of the theme. In the graphic demonstrations we will use meristic data (discrete variables) to detail the structure of the groups formed, due to the ease of formatting the examples of Figures 1–4. The results and conclusions, however, also apply to continuous variables (weights and measures), whose groupings are formed from the results of analysis of variance (ANOVA).

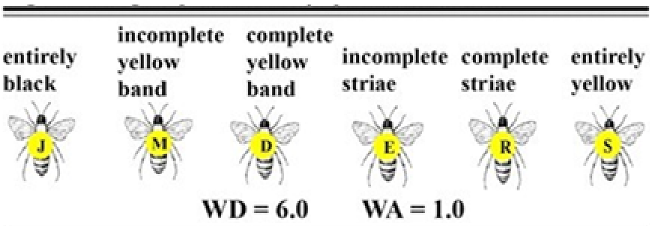

Figure 1 Subgroups formed by pronotum color.

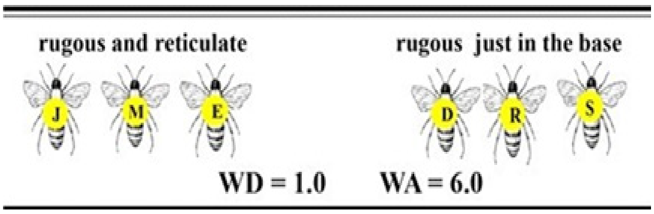

Figure 2 Subgroups formed by propodium punctuation.

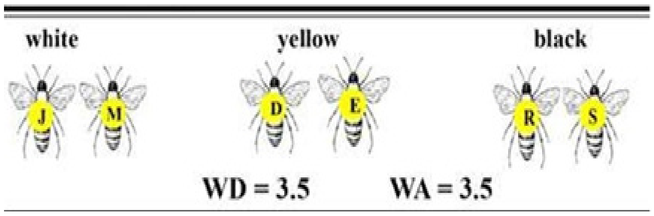

Figure 3 Subgroups formed by labrum color.

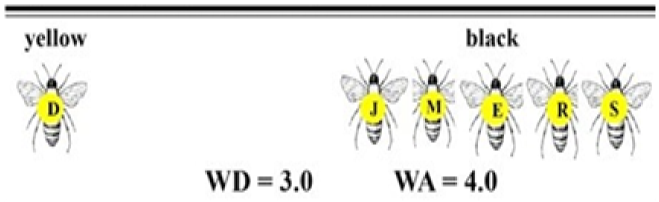

Figure 4 Subgroups formed by scapum color.

Results and discussions

The species *Plebeia juliani, Plebeia meridionalis, Plebeia droryana, Plebeia emerina, Plebeia remota and Plebeia saiqui* are also identified by the letters J, M, D, E, R and S in the figures that illustrate the presentation of the methodology. We also modified the scale of weights, which initially ranged from zero to 5, to a scale from 1 to 6. This modification consisted of increasing a unit on the original scale making it more understandable in the interpretation of the relation 'contribution of character/value of weighting. The use of weights to hierarchize the contributions of taxonomic characters is an old discussion that was potentiated with the emergence of Numerical Taxonomy in the 1950s, due to the lack of a rational and objective criterion of weighting. Michener et al.[4] advocated the use of equal weights for all characters as an alternative to work around the problem. Other solutions were presented by Burtt et al.[5] Farris et al.[6] Goodman et al.[7] Sneath et al.[8] recognized that the controversies over the use of weighting in Numerical Taxonomy were responsible for the difficulties that slowed its progress. The importance of a taxonomic character depends on the objectives of the research and its behavior, discriminating or aggregating the members of the studied group. If the study aims to individualize species, the characters capable of discriminating are particularly important. However, if the study aims to structure taxons at higher levels, formed by the aggregation of similar individuals from lower taxons, aggregating characters play a relevant role in this task. This importance can be assessed by calculating the character weights in the formations of the taxons and in the identifications of the species. Maia et al.[1] used discrete and continuous variables in the calculation of similarity coefficients using analysis of variance, without, however, creatinga weighting criterion.

Maia et al.[2] proposed the creation of a weighting scale for discrete variables (counts and encodings), formed by 5 categories, whose value of D, calculated through the formula D = 5(N – 2G) / (N-2), varies from zero to 5. For the continuous variables (measurements), Maia et al.[3] presented a formula for the calculation of D = 5(1- 2Z/(n(n-1)), where Z is equal to the numberof comparisons whose means do not differ significantly by Tukey test. In both cases the values of D vary from zero to 5, where zero represents the minimum degree of discrimination and 5 the maximum degree. Thus, twoscales of weights were formatted, one to evaluate the potential for discrimination (D) and the other to evaluate the potential for aggregation (D'). The values are complementary: D' = 5 – D. In the formatting proposed here we exclude the weight = 0, and the 5 categories have weights ranging from 1 to 6 due to the addition of a unit (+ 1)in the value of D. Thus, the weighting for discriminative characters is calculated as WD = D + 1, where WD stands for weight of the discriminative potential of the character. The weight of the aggregative potential(WA) is calculated as WA = 7 – WD. Most characters both aggregate and discriminate, that is, taxonomic characters have aggregative potential and discriminative potential, whose proportions may vary according to the case. The use of WD or WA values depends on the objectives of the study and should be done by framingthe calculated value within the corresponding limits of Table 1 & 2.

Table 1 Weight limits for weighting **(WD)** for discriminative taxonomic characters

1 <= WA <= 2	- Very little discriminative character
2 < WA <= 3	- Little discriminative character
3 < WA <= 4	- Moderate discriminative character
4 < WA <= 5	- discriminative character
5 < WA <= 6	- Very discriminative character

Table 2 Weight limits for weighting **(WA)** for aggregative taxonomic characters

1 <= WA <= 2	- Very little aggregative character
2 < WA <= 3	- Little aggregative character
3 < WA <= 4	- Moderate aggregative character
4 < WA <= 5	- Aggregative character
5 < WA <= 6	- Very aggregative character

For a better understanding of the weighting criteria, we are going to analyze how the species are organized within the studied group and the respective values of WD and WA. In Figure 1 we analyze the character 'color of the pronotum' that presents a maximum degree of discrimination. This happens when the status number of the character is equal to the number of species that make up the analyzed group. These characters are suitable for studies aimed at the individualization of taxonomic units. In Figure 2 we will analyze the 'propodium punctuation'. This character has the smallest grouping of species that a variable character can form: 2 subgroups of equal size brought together by 2 different statuses. Thus, we observed that the weight of the aggregation potential reaches the maximum value (WA = 6 .0) while the weight of discrimination reaches. The Figure 3 shows a situation in which the character 'color of the labrum' can aggregate and discriminate minimum value (WD = 1 .0). Characters with these qualities are suitable for cladistic studies, as they can form larger groups from the gathering of similar smaller groups in a similar way, forming homogeneous subgroups that differ from each other. When the WD and WA values are close, the characters are classified as intermediate (or moderate) and sometimes the WD and WA values are exactly the same as shown in Figure 3. In Figure 4 we analyze the taxonomic character 'color of the scapum' in which the aggregative potential (WA = 4.0) is greater than the discriminative potential (WD = 3.0), or, in other words, this character aggregates more than discriminates.[9]

Automation of calculations

D, WD and WA values can be easily calculated through the GADVDVC. F application, developed by the author, which includes the calculation routines for discrete and continuous variables. There are two versions, in Portuguese and in English.

Using the application

The GADVDVC. F application was developed to calculate the degree of aggregation/discrimination (D) of a taxonomic character and also to calculate the weights of a weighting system (on a scale from 1 to 6) that informs the contribution of the character to the formation of subgroups in a group (that brings together 3 or more species) (WA) or for individualization of these same species (WD). The term 'group' is being used in cases where the studied species do not represent the whole of a genus, but only part of it. Thus, the conclusions will be valid only for the group includes 16 other statistical tests of current use in biological research.

The tutorial

BIOESTAT can be obtained from the website of the Department of Statistics of UFPR, by downloading it at the address http://est.ufpr.br / 'Recursos, Software' / 'Projetos Ativos, Bioestat'). It is also important to download a tutorial extensively illustrated with the prints of the screens, exemplifying in a very didactic way the routines of the calculations through examples developed for each test studied and should not be extended to the corresponding genus. The methodological foundation for these calculations can be found in the bibliography cited at the end of this chapter.

These are two works that deal with the theme. Maia et al.[2] presents the methodology for discrete variables (counts and codifications), based on the calculation of the geometric average. Maia et al.[3] deals with the continuous variables formed by measurements using Analysis of Variance, F Test (one classification criterion), complemented by the Tukey test. In the application tutorial, yellow and green boxes are shown. The yellow boxes reproduce the application screens numbered in the sequence in which they appear during execution. The green boxes show explanatory comments about the reproduced screens. For discrete variables (counts and encodings) in use the tutorial starting on **Screen 01**. For continuous variables (measurements) use the tutorial starting on **Screen 01a (Screen 2–10) (Screen 2a–10a)**.

We will exemplify with the data of 6 species (n=6) of the genus Plebeia (*P.juliani, P.meridionalis, P.droryana, P.emerina, P.remota and P.saiqui* represented by the letters J, M, D, E, R and S, respectively). The analyzed variable is the measurement of the 'Jaw width' of 5 specimens (m = 5) of each species Maia et al.[3] (Table 3 & 4).

Table 3 Width of the jaw of six species of the genus Plebeia

J	R	S	D	E	M
0.9	1.4	1.2	1.15	1.5	1
1	1.8	1.3	1.2	1.3	0.9
1	1.4	1.5	1.15	1.2	0.9
0.9	1.45	1.3	1.2	1.2	0.7
0.9	1.4	1.3	1	1.2	0.8

Table 4 Range and studentized range tables

v/n	2	3	4	5	6	7	8	9	10
1	17.97	26.98	32.82	37.08	40.41	43.12	45.40	47.36	49.07
2	6.085	8.331	9.798	10.88	11.74	12.44	13.03	13.54	13.99
3	4.501	5.910	6.825	7.502	8.037	8,478	8.853	9.177	9.462
4	3.927	5.040	5.757	6.287	6.707	7.053	7.347	7.602	7.826
5	3.635	4.602	5.218	5.673	6.033	6.330	6.582	6.802	6.995
6	3.461	4.339	4.896	5.305	5.628	5.895	6.122	6,319	6.493
7	3.344	4.165	4.681	5.060	5.359	5.606	5.815	5.998	6.158
8	3.261	4.041	4.529	4.886	5.167	5.399	5.597	5.767	5.918
9	3.199	3.949	4.415	4.756	5.024	5.244	5.432	5.595	5.739
10	3.151	3.877	4.327	4.654	4.912	5.124	5.305	5.461	5.599
11	3.113	3.820	4.256	4.574	4.823	5.028	5.202	5.353	5.487
12	3.082	3.773	4.199	4.508	4.751	4.950	5.119	5.265	5.395
13	3.055	3.735	4.151	4.453	4.690	4.885	5.049	5.192	5.318
14	3.033	3.702	4.111	4.407	4.639	4.829	4.990	5.131	5.254
15	3.014	3.674	4.076	4.367	4.595	4.782	4.940	5.077	5.198
16	2.998	3.649	4.046	4.333	4.557	4.741	4.897	5.031	5.150
17	2.984	3.628	4.020	4.303	4.524	4.705	4.858	4.991	5.108
18	2.971	3.609	3.997	4.277	4.495	4.673	4.824	4.956	5.071
19	2.960	3.593	3.977	4.253	4.469	4.645	4.794	4.924	5.038
20	2.950	3.578	3.958	4.232	4.445	4.620	4.768	4.896	5.008
24	2.919	3.532	3.901	4.166	4.373	4.541	4.684	4.807	4.915
30	2.888	3.486	3.845	4.102	4.302	4.464	4.602	4.720	4.824
40	2.858	3.442	3.791	4.039	4.232	4.389	4.521	4.635	4.735
60	2.829	3.399	3.737	3.977	4.163	4.314	4.441	4.550	4.646
120	2.800	3.356	3.685	3.917	4.096	4.241	4.363	4.468	4.560
8	2.772	3.314	3.633	3.858	4.030	4.170	4.286	4.387	4.474

v/n	11	12	13	14	15	16	17	18	19
1	50.59	51.96	53.20	54.33	55.36	56.32	57.22	58.04	58.83
2	14.39	14.75	15.08	15.38	15.65	15.91	16.14	16.37	16.57
3	9.717	9.946	10.15	10.35	10.53	10.69	10.84	10.98	11.11
4	8.027	8.208	8.373	8.525	8.664	8.794	8.914	9.028	9.134
5	7.168	7.324	7.466	7.596	7.717	7.828	7.932	8.030	8.122
6	6.649	6.789	6.917	7.034	7.143	7.244	7.338	7.426	7.508
7	6.302	6.431	6.550	6.658	6.759	6.852	6.939	7.020	7.097
8	6.054	6.175	6.287	6.389	6.483	6.571	6.653	6.729	6.802
9	5.867	5.983	6.089	6.186	6.276	6.359	6.437	6.510	6.579
10	5.722	5.833	5.935	6.028	6.114	6.194	6.269	6.339	6.405
11	5.605	5.713	5.811	5.901	5.984	6.062	6.134	6.202	6.265
12	5.511	5.615	5.710	5.798	5.878	5.953	6.023	6.089	6.151
13	5.431	5.533	5.625	5.711	5.789	5.862	5.931	5.995	6.055
14	5.364	5.463	5.554	5.637	5.714	5.786	5.852	5.915	5.974
15	5.306	5.404	5.493	5.574	5.649	5.720	5.785	5.846	5.904
16	5.256	5.352	5.439	5.520	5.593	5.662	5.727	5.786	5.843
17	5.212	5.307	5.392	5.471	5.544	5.612	5.675	5.734	5.790
18	5.174	5.267	5.352	5.429	5.501	5.568	5.630	5.688	5.743
19	5.140	5.231	5.315	5.391	5.462	5.528	5.589	5.647	5.701
20	5.108	5.199	5.282	5.357	5.427	5.493	5.553	5.610	5.663
24	5.012	5.099	5.179	5.251	5.319	5.381	5.439	5.494	5.545
30	4.917	5.001	5.077	5.147	5.211	5.271	5.327	5.379	5.429
40	4.824	4.904	4.977	5.044	5.106	5.163	5.216	5.266	5.313
60	4.732	4.808	4.878	4.942	5.001	5.056	5.107	5.154	5.199
120	4.641	4.714	4,781	4.842	4.898	4.950	4.998	5.044	5.086
8	4.552	4.622	4.685	4.743	4.796	4.845	4.891	4.934	4.974

Table 4 Continued...

v/n	20	22	24	26	28	30	32	34	36
1	59.56	60.91	62.12	63.22	64.23	65.15	66.01	66.81	67.56
2	16.77	17.13	17.45	17.75	18.02	18.27	18.50	18.72	18.92
3	11.24	11.47	11.68	11.87	12.05	12.21	12.36	12.50	12.63
4	9.233	9.418	9.584	9.736	9.875	10.00	10.12	10.23	10.34
5	8.208	8.368	8.512	8.643	8.764	8.875	8.979	9.075	9.165
6	7.587	7.730	7.861	7.979	8.088	8.189	8.283	8.370	8.452
7	7.170	7.303	7.423	7.533	7.634	7.728	7.814	7.895	7.972
8	6.870	6.995	7.109	7.212	7.307	7.395	7.477	7.554	7.625
9	6.644	6.763	6.871	6.970	7.061	7.145	7.222	7.295	7.363
10	6.467	6.582	6.686	6.781	6.868	6.948	7.023	7.093	7.159
11	6.326	6.436	6.536	6.628	6.712	6.790	6.863	6.930	6.994
12	6.209	6.317	6.414	6.503	6.585	6.660	6.731	6.796	6.858
13	6.112	6.217	6.312	6.398	6.478	6.551	6.620	6.684	6.744
14	6.029	6.132	6.224	6.309	6.387	6.459	6.526	6.588	6.647
15	5.958	6.059	6.149	6.233	6.309	6.379	6.445	6.506	6.564
16	5.897	5.995	6.084	6.166	6.241	6.310	6.374	6.434	6.491
17	5.842	5.940	6.027	6.107	6.181	6.249	6.313	6.372	6.427
18	5.794	5.890	5.977	6.055	6.128	6.195	6.258	6.316	6.371
19	5.752	5.846	5.932	6.009	6.081	6.147	6.209	6.267	6.321
20	5.714	5.807	5.891	5.968	6.039	6.104	6.165	6.222	6.275
24	5.594	5.683	5.764	5.838	5.906	5.968	6.027	6.081	6.132
30	5.475	5.561	5.638	5.709	5.774	5.833	5.889	5.941	5.990
40	5.358	5.439	5.513	5.581	5.642	5.700	5.753	5.803	5.849
60	5.241	5.319	5.389	5.453	5.512	5.566	5.617	5.664	5.708
120	5.126	5.200	5.266	5.327	5.382	5.434	5.481	5.526	5.568
8	5.012	5.081	5.144	5.201	5.253	5.301	5.346	5.388	5.427

v/n	38	40	50	60	70	80	90	100
1	68.26	68.92	71.73	73.97	75.82	77.40	78.77	79.98
2	19.11	19.28	20.05	20.66	21.16	21.59	21.96	22.29
3	12.75	12.87	13.36	13.76	14.08	14.36	14.61	14.82
4	10.44	10.53	10.93	11.24	11.51	11.73	11.92	12.09
5	9.250	9.330	9.674	9.949	10.18	10.38	10.54	10.69
6	8.529	8.601	8.913	9.163	9.370	9,548	9,702	9.839
7	8.043	8.110	8.400	8.632	8.824	8.989	9.133	9.261
8	7.693	7.756	8.029	8.248	8.430	8.586	8.722	8.843
9	7.428	7.488	7.749	7.958	8.132	8.281	8,410	8.526
10	7.220	7.279	7.529	7.730	7.897	8.041	8.166	8.276
11	7.053	7.110	7.352	7.546	7.708	7.847	7.968	8.075
12	6.916	6.970	7.205	7.394	7.552	7.687	7.804	7.909
13	6.800	6.854	7.083	7.267	7.421	7.552	7.667	7.769
14	6.702	6.754	6.979	7.159	7.309	7.438	7.550	7.650
15	6.618	6.669	6.888	7.065	7.212	7.339	7.449	7.546
16	6.544	6.594	6.810	6.984	7.128	7.252	7.360	7.457
17	6.479	6.529	6.741	6.912	7.054	7.176	7.283	7.377
18	6.422	6.471	6.680	6.848	6,989	7.109	7,213	7,307
19	6.371	6.419	6.626	6.792	6.930	7.048	7.152	7.244
20	6.325	6.373	6.576	6.740	6.877	6.994	7.097	7.187
24	6.181	6.226	6.421	6.579	6.710	6.822	6.920	7.008
30	6.037	6.080	6.267	6.417	6.543	6.650	6.744	6.827
40	5.893	5.934	6.112	6.255	6.375	6.477	6.566	6.645
60	5.750	5.789	5.958	6.093	6.206	6.303	6.387	6.462
120	5.607	5.644	5.802	5.929	6.035	6.126	6.205	6.275
8	5.463	5.498	5.646	5.764	5,863	5,947	6.020	6.085

Citation: Maia JCS. An objective criterion for weighting taxonomic characters. *MOJ Biol Med.* 2024;9(2):67–71. DOI: 10.15406/mojbm.2024.09.00218

Conclusion

i. Every variable taxonomic character has aggregative and discriminative potentials, whose values may vary according to the number of species and their respective status that make up the analyzed group.

ii. The discriminative potential of a character can be represented numerically on a scale from zero to 5.

iii. The discriminative and aggregative potentials are complementary, that is, a 'very discriminative' character is also 'very little aggregative' and vice versa.

iv. The values of WD and WA represent the weights (participations) that the characters have in the formation of taxonomic groups and subgroups on a scale from 1 to 6.

v. WD and WA values can also be used as a criterion for prior character selection for studies employing more sophisticated or more costly methodologies.

Acknowledgments

None.

Conflicts of interest

The authors declare that there is no conflict of interest.

Funding

None.

References

1. Maia JCS. Interspecific statistical analysis of Plebeia (Apidae, Meliponi) species from Paraná (Brazil). *Acta Biol Par Curitiba.* 2017;46(1-2):39–58.
2. Maia JCS. A numerical criterion for assessing the discriminatory/aggregative potential of a taxonomic character with a Fortran 90 Program for calculations. *Acta Biol Par Curitiba*. 2021;50(1-4):7–17.
3. Maia JCS. A numerical criterion for assessing the discriminative or aggregative potentialof a taxonomic character. Part II. *Global Journal of Science Frontier Research.* 2022;22(1).
4. Michener CD, Sokal RR. A quantitative approach to a problem in classification. *Evolution.* 1957;11(2):130–162
5. Burtt BL. Angiosperm taxonomy inpractise. *Phenetic and Phylogenetic Classification.* Sys Ass Pub. 1964;6(164):5–16.
6. Farris JS. Estimation of conservatism of character by constancy with biological populations. *Evolution.* 1966;20(4):587–591.
7. Goodman MM. Measuring evolutionary divergence. *Jap J Genet*. 1969;44(suppl. 1):310–316.
8. Sneath PHA, Sokal RR. Numerical taxonomy. The principles and practice of numerical classification. *CABI Digital library.* 1973;573.
9. Harter H, LEON. Tables of range and studentized range. *Ann Math Statist.* 1960;1(4):1122–1147.