COMPARISON OF DIFFERENT TAXONOMY PROCEDURES

Dobromir Bonacin
Faculty of kinesiology, Travnik, Bosnia & Herzegovina

Review paper

Abstract
The purpose of this study was the comparison of the different taxonomy procedures which are universally applicable to social studies. The design of the study was pointed towards the identification of the characteristics of the individual taxonomy procedures with the algorithmic coding in Borland Delphi developing tool (D.Bonacin). The data of 249 students described with 14 morphological variables was processed for the purpose of this study. The results have shown the divergences in different taxonomy procedures. It was possible to identify the logic of these procedures, and to suggest the application of different models in particular scientific situations. Methodological constraints of the study are possible to find only if a more complex algorithm which entails all of the characteristics of the analyzed procedures is generated. The originality of the study lies in the completely new approach to the object classification, complete with the proposition which procedure to apply in different situations.

***Key words**: classification, taxons, polar, distinct*

Introduction

In every scientific discipline there is a classification of the objects which are studied. The term classification basically corresponds to the formation of groups which are not known to the arbitrarily defined total sample, described with an arbitrary number of quantity variables. In that way the groups which are by some criteria scientifically sustainable are extracted from the total mass. As we can see, the idea here is the identification of the unknown phenomena. Different forms of classification are long known, but their character was mostly arbitrary, because the final result was much dependant on the settings of the scientist involved, and that made the final outcome very disputable. The situation deteriorated with the introduction of the even greater number of variables with which the objects are described.

However, with the development of the computers, there followed a great number of algorithms and programs for the object classification. Although these procedures cannot are not without merit in some situations, we must point out that the problem of classification (taxonomisation) is not solved up to date. In scientific practice there is still a doubt connected to the taxonomisation, a doubt which stems from serious contemplations affiliated with distribution of the human traits.

According to some, these traits are usually normally distributed which means that it is not realistic to expect the real existence of real taxons (types) inside the samples represented , and in this case the taxons are only artifacts. On the other hand, many research shows that in large number of human dimensions (morphological, cognitive... etc) there is initially no multivariable normal distribution, but a distinct and multi-modal distribution, so the existence of the real taxons is not a mere mathematical artifact.

For these reasons exactly, we need to approach these taxonomy procedures with a lot of attention and with respect for the real knowledge of the analyzed phenomena.

Basic taxonomy models

For the purpose of this work the examples from three areas will be analyzed: a) Distinct taxons (the taxons which allow the allocation of one object on one, and only one taxon), b) Polar taxons (the taxons which take to extremes one latent taxonomic dimension on the opposite side of latency , and which allow non-trivial allocation of one object on more than one object), and c) The alternative model which entails good traits of both of the above mentioned models.

The distinct model - Hierarchical grouping

This procedure basically tries to group the object in order to enable the hierarchical allocation into the smallest group based on the successive adjoining (Veldman, 1967). The two big flaws of this procedure are that there are no levels, but multidimensional adjoining on the basis on similarity, and the objective number of really existent taxons.

The Distinct Model - The Uditax extremes

This procedure (Bonacin, 2004) proved itself as one of the best procedures in over a 100 real situations, even when the groups of objects were vaguely defined. The procedure initially defines extreme objects (initial micro-taxons) and later one by one taxon is added, creating the real taxonomic groups. Analytically, this is probably the best taxonomic procedure.

The alternative model - integrating taxons

This procedure (Bonacin, 2002) is the upgraded model of the hierarchical grouping and it produces additional information and objectifies the number and the structure of the taxons derived by some other method. The results allow fairly objective conclusions and interpretations.

Poltax model of the polar taxons

This procedure (ex. Momirović et al, 1987) is a generalization of the Catell logic by which the taxonomic dimensions do not have their own projections in relation to centroids of the groups, but in relation to extreme groups of objects, and so one taxonomic dimension has a dual role and one common trait. This logic was proven as equally valid and equally bad, which led to the series of doubts in terms of adequacy.

The example of the results

For the comparison of the results' quality of taxonomic analysis, 249 male entities aged 7 (+/- 2 months) have been analyzed using 14 quantitative variables designed to completely cover morphological area, latently dimensioned as: a) longitudinality (height, arms' length, legs' length), b) transversal (knee diameter, wrist diameter, biacromial range, bicristal range), c) body mass and volume (weight, forearm diameter, calf diameter, chest diameter, d) fat tissue (skinfold on forearm, skinfold on the back, skin fold of the belly). All the variables are standardized to level the influence of individual variables. This example is chosen because it is rather straightforward, although much more complex cases have been compared.

The Distinct Model - Hierarchical Grouping

G	GR01	GR02
N	194	55
%	77.91	22.09
AVIT	-0.22	0.77
ADUN	-0.20	0.72
ADUR	-0.21	0.76
ADRZ	-0.20	0.69
ADIK	-0.22	0.79
ASIR	-0.22	0.76
ASIK	-0.23	0.82
ATEZ	-0.40	1.41
AOPL	-0.31	1.09
AOPK	-0.34	1.21
AOGK	-0.35	1.22
AKNN	-0.36	1.28
AKNL	-0.33	1.17
AKNT	-0.39	1.37

The model of hierarchical grouping has shown as a result two groups which are produced in accordance to the criteria of growth of the total grouping error.

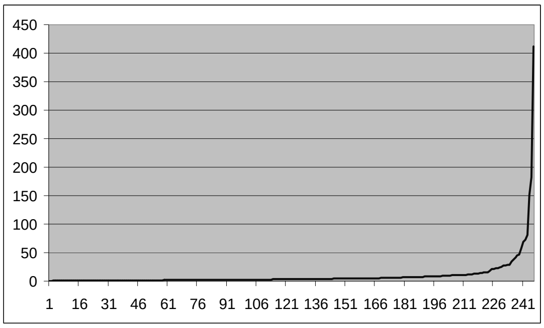

Figure 1. Group error in simultaneous phases

In the first group there are 194, and in the second 55 objects. The analysis of the dendogram has shown that the grouping was not convincing enough, and neither the number nor the character of the given taxons is certain. The average values of the taxons on the variables show that the first taxon defines completely average entities, but with the values slightly under the average. In other words, the objects with medium or small body mass, small volume, low longitudinal, and with little body fat. The second one, on the opposite, defines the objects with much bigger body mass, especially fat tissue and volume.

The Alternative Model - Integrating Taxons

	PRX1	PRX2
AVIT	0.41	-0.03
ADUN	0.35	0.01
ADUR	0.37	0.01
ADRZ	0.26	0.08
ADIK	0.32	0.07
ASIR	0.46	-0.08
ASIK	0.27	0.13
ATEZ	0.21	-0.44
AOPL	0.24	-0.28
AOPK	0.22	-0.34
AOGK	0.20	-0.36
AKNN	-0.05	-0.62
AKNL	-0.10	-0.61
AKNT	-0.10	-0.70
	PRX1	PRX2
PRX1	1	0.33
PRX2		1

The model of the integrating taxons, out of which the data is brought to diagonally rotating Promax taxons on the variables, has shown that two isolated taxons are much clearer than in the previous example. Namely, the first one describes the objects with the defined skeleton dimensions, while the second one obviously is growth and soft tissue development mechanism (musculature and soft tissue).

It is still 194+55 objects, but their definition is much more precise in the space of variables. It is also visible that these two taxons are not orthogonal, but have very positive relations (r=0,33), the number derived from the dimensions of volume, as to say clearly non-differentiated morphological dimensions at this age. As we see, the alternative model of the taxon has shown a more precise definition than the hierarchical grouping.

The Distinct Model-Uditax positions on the basis of the extreme objects

The results of the Uditax taxonomic procedure have shown the existence of the two well defined taxons, where the first certainly entails mass, volume and skeleton measurements. All of the morphological indicators are equal in the range of average value from 0.50 to 0.80.

The other taxon is constituted in the identical way, but with the negative value, with which it forms the general typology of children very similar to hierarchical grouping, but with more prominent values.

It is noticeable, however, that the number of objects is significantly different, there is now 114, or 135 objects per taxon.

	DT01	DT02
N	114	135
%	45.78	54.22
AVIT	0.64	-0.54
ADUN	0.56	-0.48
ADUR	0.57	-0.48
ADRZ	0.64	-0.54
ADIK	0.64	-0.54
ASIR	0.55	-0.47
ASIK	0.54	-0.46
ATEZ	0.82	-0.69
AOPL	0.68	-0.57
AOPK	0.73	-0.62
AOGK	0.73	-0.62
AKNN	0.55	-0.47
AKNL	0.47	-0.39
AKNT	0.58	-0.49
Positions of the		
	DT01	DT02
GR01	2.35	-1.99
GR02	-2.35	1.99

This describes the real situation much more accurately (on the basis of the insight on the initial data). It is noticeable that the Uditax procedure has integrated the object with large fat tissue and those with big volume, which is justified considering the fact that the morphological differentiation has not yet occurred at this age. The relation of these two taxons is diametrically opposite, which is logical, because we highlight two divergent developmental attributes.

The Polar Model- Polar Taxons Model

	TX1	TX2
AVIT	-0.11	0.97
ADUN	-0.11	0.90
ADUR	-0.12	0.91
ADRZ	0.14	0.62
ADIK	0.23	0.58
ASIR	0.16	0.62
ASIK	0.34	0.40
ATEZ	0.65	0.43
AOPL	0.67	0.19
AOPK	0.65	0.28
AOGK	0.74	0.18
AKNN	0.97	-0.22
AKNL	1.05	-0.39
AKNT	1.01	-0.25
Correlations of the taxons		
	1	0.55
		1
% of the objects		
	53.71	46.29

Discussion and Conclusion

The results of the polar taxons' analysis have shown that it is possible to push to extreme the data in the way that the first taxon represents the development of the soft tissue (fat tissue dominant), while the other represents skeleton growth (longitudinality dominant). The other dimensions (volume, mass) are present to a certain degree in both taxons, meaning that they do not fully contribute to their real typological definition. The two mentioned mechanisms of morphological growth and development are in medium and significant correlations (r=0.55), and those relations are predominant in this analysis. The percentage of the entity allocation should be regarded with a reserve, but the number of the objects is relatively balanced (53% and 46%, 134 and 115 objects). Taxonomic procedures have shown that, although the final results are fairly similar, there are significant deviations , so the question which procedure to trust is justified. It is obvious that it cannot be clustering by hierarchical grouping, and even though the alternative model has brought some improvement, the number of objects per taxon is still not balanced, which after a detailed check of the initial data is not acceptable. The Polar taxons and the Uditax have brought similar results considering the number of objects, but there are differences in accordance with the basic principles of the two procedures. Namely, while Uditax is a provider for the explication of distinct taxons, Poltax is a polar model. It seems justified to assume that the extremization is justified because it tends to extract real groups ina pattern, and both procedures are based on extremization. It is noticeable that the integrating taxons 'act' similarly to polar taxons, and Uditax similarly to hierarchical grouping. We can conclude that there is two-dimensional matrix which helps to decide what procedure gives objective results. In one dimension there is a number of objects, and in other distinct/non-distinct orientation.

Table 1. The logic of the two-dimensional matrix for deciding

	Non-balanced	Balanced
Distinct	Hierarchical group.	Uditax
Polar	Integrating	Poltax

It is safe to recommend all four procedures in the following cases: a) Hierarchical grouping with distinct taxons with non-balanced number of objects per taxon, b) Integrating taxonomy model with polar taxons with non-balanced number of objects per taxon, c) Uditax taxonomy model with distinct taxons with balanced number of objects per taxon, and d) Poltax taxonomy model with polar taxons with balanced number of objects per taxon. It is generally possible to recommend Uditax and Poltax models.

References

Bonacin, D. (2002). *Integrativni taksonomski model kao preduvjet programiranja rada u kineziologiji. Zbornik radova 11. ljetne škole kineziologa R.Hrvatske, Rovinj, 103-104.*

Bonacin, D. (2004). *Uvod u Kvantitativne metode. Vlastito izdanje.*

Bonacin, D., & Carev, Z. (2002). *The universal methodology of process identification. Journal of Theoretics. Vol.: 4, 2.(http//www.journaloftheoreitcs. com/Links/links-papers.htm).*

Bonacin, D., & Carev, Z. (2002). *Process identification. Kinesiology – new perspectives, III international scientific conference, Opatija, Proceedings: 632-635.*

Harman, H.H. (1970). *Modern Factor Analysis.* Chicago: The University of Chicago.

Hošek, A. (1981). Povezanost morfoloških taksona s manifestnim i latentnim dimenzijama koordinacije. *Kineziologija, 11*(4): 5-108.

Johnson, R.A., & Wichern, D.W. (1992). *Applied multivariate statistical analysis*. Englewood Cliffs: Prentice-Hall.

Momirović, K, Prot. F., Dugić, D., Knezović, Z., Bosnar, K., Erjavec, N., Gredelj, M., Kern, J., Dobrić, & V., Radaković, J. (1987). *Metode, algoritmi i programi za analizu kvantitativnih i kvalitativnih promjena. Institut za kineziologiju FFK Sveučilišta u Zagrebu.*

Momirović, K. (1984). *Kvantitativne metode za programiranje i kontrolu treninga. Statističke metode 1.* Zagreb: FFK.

Mulaik, S.A. (1972). *The foundations of factor analysis.* McGraw-Hill, New York.

Veldman, D.J. (1967). *Fortran programming for behavioral sciences*. New York: Holt, Rinehart & Winston.

POREĐENJE RAZLIČITIH TAKSONOMSKIH PROCEDURA

Sažetak

Svrha ovog istraživanja bila je poređenje različitih taksonomskih procedura koje je moguće standardno koristiti u društvenim naukama. Dizajn istraživanja bio je usmjeren prema identifikaciji karakteristika pojedinih taksonomskih procedura uz kodiranje algoritama u Borland Delphi razvojnom alatu (D.Bonacin). Za potrebe ovog istraživanja i primjer obrađeni su podaci 249 učenika opisanih sa 14 morfoloških varijabli. Rezultati su pokazali da postoje divergencije u dobivenim rezultatima različitih taksonomskih postupaka. Bilo je moguće identifikovati logiku tih postupaka i kao praktične implikacije predložiti primjenu pojedinih modela u karakterističnim naučnim situacijama. Metodološka ograničenja istraživanja moguće je pronaći samo ukoliko se generira neki složeniji algoritam koji bi objedinio sva navedena svojstva analiziranih postupaka. Originalnost rada ogleda se u posve novom cjelovitom pristupu klasifikaciji objekata s prijedlogom primjene postupaka u kartakterističnim situacijama.

Ključne riječi*: klasifikacija, taksoni, polarni, distinktni*

Received: September, 11. 2008.
Accepted: December, 10. 2008.
Correspondence to:
Assist.Prof. Dobromir Bonacin, PhD.
University of Travnik
Faculty of kinesiology
Aleja konzula 5, Travnik, BiH
Phone: +387(0)30 517 824
E-mail: dobromirb@sposci.com