

UDITAX – MODEL OF DISTINCT TAXONS BASED ON DISTANCE MAXIMIZATION OF REPRESENTATIVE ENTITIES

Dobromir Bonacin¹, Danijela Bonacin¹ and Žarko Bilić²

¹ Faculty of Education, Travnik, Bosnia & Herzegovina

² Faculty of Sciences, Mostar, Bosnia & Herzegovina

Original scientific paper

Abstract

The aim of this research was to define algorithm and coding of the new model of taxon analysis based on representative entity maximization in total sample. Methodological design is oriented toward new approach to distinct taxons that enables representatives in total sample and they become carriers of primary group information. With remaining entity successive accession practical groups, sub-samples i.e. taxons are being formed. The results presented real existence of taxons in a few situations and for illustration we selected complex solution with 249 boy's age of 7 described with 26 biomotorical variables (morphological, motor and functional). Originality of this research and presented model is placed in stability of extracted taxons and their extremely simple identification. From a strictly mathematical point of view this model of taxon analysis is totally impassive to the number of entities or variables and as such is applicable on small samples that are usually used in sports. Practically there are no limitations of a model and presented algorithm so it is even possible that the amount of variables is multiple higher then the amount of entities which usually presents limitations for every other model.

Key words: grouping, distances, representatives

Introduction

Even though taxonomic procedures are long above project and laboratory environment, they are not sufficiently applied in daily practical work. (Veldman, 1967; Momirović, 1987; Johnson & Wichern, 1992). One of the causes for such situation is insufficient education of suitable personnel but we must not neglect unnecessary complications in their generation and representation (Bonacin, 2003; Bonacin, 2004). Also, in literature, we often find not very logical opinion indicates existence of polar taxons only. This speculative opinion is explained with disability to divide the set of multivariate normally distributed variables into sub-groups that would be sufficiently distant from one another. This is incorrect, since in every, no matter how defined hyperellipsoid (hyper ball is the exception), exist at least two taxons that actually form such ellipsoid and "extract" it by longer axis. (Fulgosi, 1979). If we monitor them under distinct model we will obtain two extreme sets which on their ends have their obvious representatives. By logical generalization of this model it is possible to select arbitrary number of entity sets under assumption that it is possible to find two or more extreme entities in multidimensional space. Such procedure would certainly have to be totally resistant to small number of objects or variables.

Algorithm

Lets say that from population E we have random object sample e_i ($i=1..x$) described with set of variables v_j ($j=1..y$) extracted from population V. With Hadamard's value accession to objects we obtain gross data matrix $B = E \otimes V$ with elements $b_{i,j}$ ($i=1..x, j=1..y$). In such defined space, the distance between objects is determined as $c_{i,j} = [\sum(b_i-b_j)^2]^{1/2}$, where C is square and symmetric matrix of Euclidean distances of objects with elements $c_{i,j}$ ($i=1..x, j=1..x$). With spectral decomposition of this matrix we obtain specific vectors X with elements $x_{i,j}$ ($i=1..x, j=1..z$), where z is number of nontrivial specific matrix value C.

This number can be determined so it equals PB criteria, i.e. total mutual variance, then to the number of extraordinary eigen values, also to pre-default value or to scree technique equals to the last nontangential value of all specific values sorted in order of falling values. Retained characteristic vectors X are rotated into orthogonal parsimonious position to meet the function

$$m \sum_{i=1}^m \sum_{j=1}^r (f_{ij}/h_i)^4 - \sum_{j=1}^r \left(\sum_{i=1}^m f_{ij}^2/h_i \right)^2 = \max$$

known as Kaiser's normalized Varimax criteria.

After that non-orthogonal parsimonious position is maximized e.g. Othoblique or unconditioned Promax. Based on projections of oblique-angular set A with elements $a_{i,j}$ ($i=1..x$, $j=1..z$), in every column we discover maximal absolute value $\text{abs}(\max a_j)$ ($j=1..z$), which defines representative for each taxon. In this way we certainly achieved that taxon representatives are maximally distant from one another for a given number of taxons z.

Mutual distance of representative objects is defined with matrix G with elements $g_{i,j} = [\sum(g_i - g_j)^2]^{1/2}$ ($i=1..z$, $j=1..z$), and the distance of remaining objects from matrix C with elements $c_{i,j}$ $[\sum(c_i - c_j)^2]^{1/2}$ ($i=1..x-z$, $j=1..x-z$) is defined with matrix U with elements $u_{i,j} = [\sum(c_i - g_j)^2]^{1/2}$ ($i=1..x-z$, $j=1..z$). In the following actions successively one per one object is being joined with one of the existing taxons G based on the least distance in the matrix U. Taxon centroids in original and standardized matrix result with precise information about object affiliation to taxons, as well as position of the groups in variables from set V.

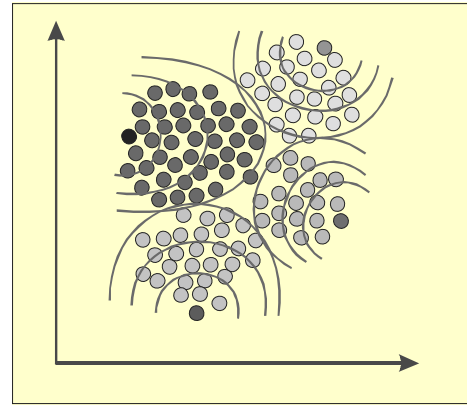


Figure 3. Uditax iterative

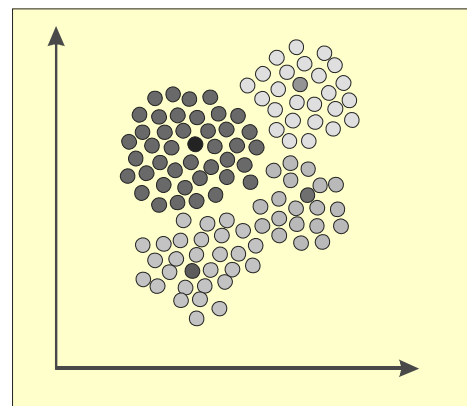


Figure 4. Uditax final solution

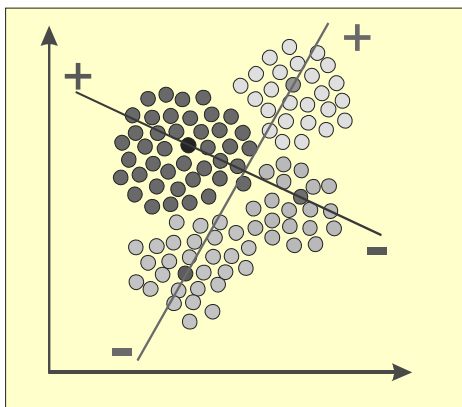


Figure 1. Polar taxons solution

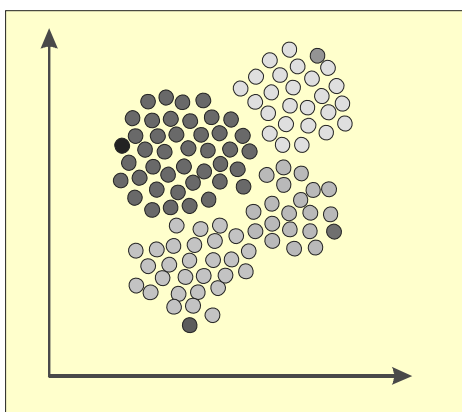


Figure 2. Uditax Representatives

Naturally, procedure obtains many results, but usually we present (table 1) only basic. Algorithm defined, coded and tested by D. Bonacin, PhD.

Numerical sample

Group of 249 boys age of 7 was subject to 14 morphological variable testing (*body height – AVIT, leg length – ADUN, arm length – ADUR, wrist diameter – ADRZ, knee diameter – ADIK, biacromial width – ASIR, bicristal width – ASIK, body weight – ATEZ, forearm circumference – AOPL, lower leg circumference – AOPK, chest circumference – AOGK, upper arm skin fold – AKNN, skin fold of the back – AKNL, abdominal skin fold – AKNT*). Also we applied 11 motor variables (*side steps – MKUS, polygon backwards – MPOL, standing on the bench – MP20, straddle forward band – MPRR, hand tapping – MTAP, foot tapping – MTAN, long jump from a stand still – MSDM, throwing ball from a distance – MBLD, 20 m run from a standing start – M20V, sit-ups – MDTS, held part in the hang – MVIS*) and one functional variable 3 minute run – MT3M). For the purpose of selecting the most capable ones for sport school program, the data were processed with UDITAX – algorithm.

Table 1. Middle values and taxon centroids

	DT01	DT02	DT03	DT04	DT05
N	35	5	3	80	126
%	14.06	2.01	1.20	32.13	50.60
AVIT	-1.33	0.73	-0.21	0.75	-0.13
ADUN	-1.23	0.85	-0.34	0.68	-0.12
ADUR	-1.09	0.23	-0.08	0.69	-0.15
ADRZ	-1.07	0.82	0.73	0.60	-0.13
ADIK	-0.94	1.72	0.83	0.65	-0.23
ASIR	-1.18	1.26	1.40	0.62	-0.14
ASIK	-0.99	1.92	1.12	0.58	-0.19
ATEZ	-1.12	2.78	0.15	0.92	-0.39
AOPL	-0.95	1.79	0.35	0.67	-0.25
AOPK	-0.99	2.28	1.22	0.74	-0.31
AOGK	-0.95	2.61	0.60	0.74	-0.32
AKNN	-0.42	2.92	-0.66	0.68	-0.44
AKNL	-0.30	4.18	-0.52	0.46	-0.37
AKNT	-0.53	3.16	-0.42	0.69	-0.41
MKUS	0.58	0.48	-2.03	0.14	-0.24
MPOL	0.39	0.38	-1.18	0.29	-0.28
MP2O	0.08	0.09	1.53	-0.11	0.03
MPPR	-0.02	-0.32	0.84	0.21	-0.13
MTAP	-0.06	0.51	-0.21	-0.22	0.14
MTAN	0.00	0.76	-0.11	-0.20	0.10
MSDM	-0.61	-0.48	2.42	-0.42	0.41
MBLD	-0.73	0.25	3.34	-0.08	0.20
M20V	0.45	0.75	-1.17	0.31	-0.26
MDTS	-0.65	0.40	0.67	-0.13	0.24
MVIS	-0.21	-0.69	2.43	-0.42	0.31
MT3M	-0.22	-0.22	1.27	-0.16	0.17

Data obtained were really simple for interpretation. As seen from the results, algorithm recognized 5 groups of students and divided them in the way to maximize their abilities and characteristics. For the first taxon (14,06% of the entity) extremely small height and mass with the lack of fat tissue and extremely weak motor abilities in whole spectrum is characteristic. We probably have the case of undeveloped children regarding their age. On the other taxon is smaller group (2% of the entity) with expressed height and mass, especially fat tissue and with poor motor output except frequency in general, which is already known situation in kinesiology because it obviously describes adiposity respondents. The third taxon out of 1.20% of entity presents students of barely average mass and height, voluminous and extremely good motor abilities in every way. This group is the most capable of all. The forth taxon with 32.13 % of the entity presents morphologically well developed children but who show poor motor skills.

Literature

- Bonacin, D. (2003). Integrativni taxonomski model kao preduvjet programiranja rada u kineziologiji. [*Integrative taxonomic model as precondition for programming of work in kinesiology*]. /In Croatian/, *Ljetna škola kineziologa RH*, Rovinj, Zbornik radova : 103-104.
- Bonacin, D. (2004). *Uvod u kvantitativne metode*. [*Introduction to quantity methods*]. /In Croatian/, Kaštela: Vlastito izdanje.
- Fulgosi, A. (1979). *Faktorska analiza*. [*Factor analysis*], /In Croatian/, Zagreb: Školska knjiga.
- Johnson, R. A., & Wichern, D. W. (1992). *Applied multivariate statistical analysis*. Prentice-Hall.
- Momirović, K. (1987). *Metode, algoritmi i programi za analizu kvantitativnih i kvalitativnih promjena*. [*Methods, algorithms and programmes for analysis of quantitative and qualitative changes*], /In Croatian/, Zagreb: Fakultet za fizičku kulturu.
- Veldman, D, J. (1967). *Fortran programming for the behavioral sciences*. New York: Holt, Rinehart & Winston.

Here we probably have the case of the children who did not finish integrated level of motor. And finally the last fifth taxon with the most (50.60 %) of children describes morphological set with extraordinary entities in the sample with good motor characteristics along the whole solution. This part of population is usually the most visible. It is obvious that the set of taxons precisely described solutions (types) within defined sample and recognition of biomotor types was conducted in a very simple and easily interpreted way. We found very significant that extreme individuals, that can be valuable for sport selection procedures, were isolated. With inspecting individual basic results of entities, these results are fully confirmed.

Conclusion

Because of well known limitations of taxonomy procedures required for research projects in kinesiology, sociology, medicine, psychology and other scientific disciplines, the new model of distinct taxons was defined. It was established on representative entities that algorithm discovers and locates regarding groups that they potentially belong to. After algorithm identifies it it is joined with entities one by one, and it forms really existent sub-groups (taxons) within initial sample. Algorithm has no limitations that could result from the number of variables or entities and as such it is ranked on the top of such procedures for research projects. Quality of procedures was tested in a few data structures in various situations (from a very clear types to unfavorable set of data) to obtain information on quality and abilities. For this approach we selected complex sample with 249 boy's age of 7 described with 26 biomotor variables. Identified taxons had clear and recognizable configuration. These results were confirmed with detail inspection of initial data. This procedure is recommended especially when it comes to sport groups that usually have problems with levels of freedom and representation.

UDITAX – MODEL DISTINKTNIH TAKSONA TEMELJEN NA MAKSIMIZACIJI UDALJENOSTI REPREZENTATIVNIH ENTITETA

Sažetak

Svrha ovog rada bila je definiranje algoritma i kodiranje novog modela taksonomske analize koji je utemeljen na maksimizaciji reprezentativnih entiteta u ukupnom uzorku. Metodološki dizajn je orijentiran prema novom pristupu distinktnim taksonima koji omogućava identifikaciju reprezentanata u totalnom uzorku koji tako postaju nositelji primarnih grupnih informacija. Sukcesivnim pridruživanjem ostalih entiteta formiraju se realne grupe kao sub-uzorci, tj. taksoni. Rezultati su pokazali stvarnu egzistenciju taksona u više situacija od kojih je za ilustraciju izabrana vrlo kompleksna solucija sa 249 dječaka uzrasta 7 godina opisanih sa 26 biomotoričkih varijabli (morfološke, motoričke, funkcionalne). Originalnost istraživanja i ponuđenog modela je u stabilnosti ekstrahiranih taksona i njihovoj izuzetno jednostavnoj identifikaciji. Striktno matematički gledano ovaj model taksonomske analize potpuno je neosjetljiv na broj entiteta ili varijabli, te je baš zato primjenjiv u malim uzorcima kakvi su najčešće u sportu. Ograničenja modela i ponuđenog algoritma praktično ne postoje, te je čak moguće da je broj varijabli višestruko veći od broja entiteta, što je uobičajeno ograničenje svih drugih modela.

Ključne riječi: grupiranje, udaljenosti, reprezentanti

Received: February 04, 2008.

Accepted: May 16. 2008.

Correspondence to:

Assist.Prof.Dobromir Bonacin, Ph.D.

University of Travnik

Faculty of Education

Aleja Konzula 5, 72270 Travnik, B&H

Phone: +387 (0)62 907 430

E-mail: dobromirb@actakin.com