EXAMPLE BEHAVIOR OF ALGORITHM AND PROGRAM FOR QUANTITATIVE AND STRUCTURAL MULTIVARIATE DIFFERENTIATION OF GROUP OF ENTITIES WITH DIFFERENT MORPHOLOGICAL STATUS

Dobromir Bonacin and Danijela Bonacin

Faculty of Education, University of Travnik, Bosnia and Herzegovina

Original scientific paper

Abstract

The purpose of this work was the preparation and testing of algorithm prepared for the analysis of structural multivariate distinguishing groups of entities derived from the total distributed deployments. The fundamental methodological pattern is evident in the fact that within the sample defined in any way that there are sub-samples on the basis of some objective criteria can be classified at least as inferior, average and superior, or at transformational process as initial, transitive and final state. The algorithm has been tested on several examples of which two are prepared as example. One is presented in this paper as an example of stable subsamples in morphological domain that do not show differences in the structure of the transformation process. Another example was published in the "Sport Science" journal in the motor domain and shows exactly drastic structural changes how they are, just the algorithm can detect (Bonacin & Bonacin, 2012).

***Key words**: algorithm, groups of entities, morphology, structure, distinguishing*

Introduction

Multivariate analysis, of course, are the only procedures that can give an objective picture of the world in which we are a part and to ensure convergence of the natural laws of science (Harman, 1970; Cooley & Lohnes, 1971; Mulaik, 1972, Johnson & Wichern, 1992; Momirović et al., 1987). That's why the solution to the problem of determining the structural differences in the patterns that are predefined as potentially significant structurally different is a serious problem in all disciplines in which there is unity of the dynamic features of objects, transformations, and long-term effects (Bonacin & Smajlović, 2005; Bonacin et al., 2008; Bonacin et al., 2012). Such processes are of course all the processes of physical education, all educational, all medical, and also all in which there is a time delay of reaching effects, such as economic or management.

Since the objects (egg, people in our examples) differ in the anthropological features, including the ability for adaptation or reservoirs that absorb incoming information and tasks, then it is quite likely that the consequences of applying the operator to be different, and we should not forget the fact that differences exists often already in the start of groups or profiles, because of sufficiently different objects. For these reasons it is of multiple interests to explore, and systematically explore the diversity of sub-groups of the population, mainly in relation to that very group they belong to all subgroups. Although this can be done by taxonomic analysis, the fact remains that such an analysis, it is clear, even in close coordination with the discrimination analysis, however, remain in a quantitative field of preferred phenomena explication, but it is very difficult to apply in the analysis of changes in the structure that is in qualitative area.

For these reasons, it is prepared, tested and implemented algorithm "SubStru" in order to provide information on the structural differences between the subsamples.

Algorithm

If $E = (e_i ; i=1,...,n)$ is set of entities randomly selected from some population **P** and $V = (v_j ;j:=1,...,)$ is set of linearly independents, normally distributed quantitative variables. Then with operation of joining values from **V** with entities from **E** the result is matrix $X = E \otimes V$ which explains state of set **E** on the set **V** at some point of time. If $M= s(m_j ; j=1,..m)$ is vector of middle values on m variable of matrix **X** ($m_j= \Sigma x_i / n$) and vector $S = (s_i ; j=1,..m)$ contains standard variable variations from **V** presented in **X** ($s_j = sqrt(\Sigma (xi - X)^2)/n$) then in matrix **Z** ($z_{ij} ; i=1,..n, j=1,...,m$), we will find standardized entity result values per each variable expressed in values of standard deviations of each variable ($z_{ij} = (x_{i,j} - m_j)/s_j$).

Let $E = (e_i , i = 1, ... , n)$ is some set of entities in general, with some procedure defined as sample from some population **P**. Let we have defined set of variables $V = (v_j , j = 1, ... , m)$. Let suppose that effective n of sample E is big enough, so m<<n.

Let

$$Z = E \otimes P \mid Z^t e = 0 , dg (R) = I$$

where e is summation vector (n,1), then

$$R= Z^t Z$$

is correlation matrix of results, and **Z** is matrix of standardized results in variables from **V**.

Let there exists a set of conditions that ensures independent trader objective assessment of the success set of entities in an some activity or activity which properties can be described with a several (m) linearly independent variables of **V**, which allocates an arbitrary number of entities in the group k (k = 3 standard, w_i, i= 1, .. k) but the number k can be get by taxonomisation of total effectives from **P** or with regression coefficient *Beta* with screening technique of saliency. *Beta* coefficients are then when the criterion is undoubtedly some objective indication that score or placement. The effective overall then decomposed into k subsets, one of which is extremely superior, one extremely inferior and the rest are located within these extremes, while not necessary for the subsets with the same number of elements although it is desirable to have an estimate. Under the criteria of the highest credibility then for each of the subsets, we can calculate the local association in the set of variables

$$R_w = Z_w^t \, Z_w$$

ie, for example, k = 3, then R_1, R_2 and R_3 are the correlations between variables in each of the subsets of entities from **E**.

Given that in these subsets is a virtually the same set **E**, it is realistic to assume that the overall multivariate measure of consistency of each of the samples to **E** is defined with some function of common association. Such associations are, for example, correlation, and **E** in **R** so then **Y** = **R**, but it could be also a simple additive function like average correlation of the R_1, R_2, ..., R_w or **Y** = avg(R_1 + R_2 + ... + R_w).
Now, operation

$$A_w = R_w - Y$$

results in differences in the local association (correlation) for each subset of the entities, and in the matrix

$$G_w = A_w * A_w^T$$

it will be virtual covariance that describe the differences of each association in relation to a subset of the total set of entities defined from b. This matrix can be considered a realistic description of the structural differences of particular sets of entities in relation to the total of the effective entity from **E**. Of course, assuming that in the total sample exist same set of latent mechanisms that will surely be worth it as a subset of **w** closer to average or otherwise defined common status of the total sample, and it will be the difference in A_w smaller, respectively. This hypothesis is of course possible to test by some of the other methods (Bonacin, 2010), and this principle works before applying this algorithm, but it is not the subject of this method, just as it is possible that the whole system could be transferred to some other area, such as Guttman's image or Bonacin's Iterim and then it could be possible to apply all other procedures (Bonacin, 2006).

Testing of the difference significance we can provide in two directions. One is to test the difference between the subsets, where such tests are k(k-1)/2, but usually for a small number of subsets (k <= 5), such a test may not have much meaning for the global assessment of the difference since the subsets and defined based on the different achievements, and can be used for detecting global statistical significance or impact of individual variables from **V** to form a difference. Such a test, however, certainly makes sense in another case for which this algorithm is made, and it is for overall statistical significance of the difference of multivariate associations between subsets (R_w) and general measures in the sample (**Y**), regardless of how these general measure (**Y**) was obtained.

Significance of possible obtained differences for different subsets of G_w can be easily tested, since at least one function of trace of matrix G_w

$$g = (n/(m-1)) * Trace(G_w)$$

and finaly   f = g * g

has a chi-square distribution with degrees of freedom which are a function of the number of variables in **V** so df = m.

Furthermore, the algorithm can be applied in a situation where one and the same group of entities undergoing a transformation process in which the expected changes in the structure of the dimensions are affected (education, training, ...). Then, simply states in which estimates are made for certain successive state are just data in R_w. Of course, the algorithm makes factor structure of the total sample and subsamples using principal components model with keeping the number of significant components in accordance with the PB criterion and finally rotation in orthoblique position. On special request can be thus obtained latent dimensions analyzed in terms of the subsamples.

The algorithm is designed, tested, coded and implemented in August 2012. Program adaptation for computational device is made by Dobromir Bonacin with the software tool "Delphi 2009" (Delphi 12, code named Tiburón).

Results – Numerical example

Therefore it is interesting to study the behavior of this model in reality and what differences will be between subgroups. To verify quality of algorithm we analyzed data of 249 male entities, all just turned 7 +/- 2 months, Elementary school First grade students who were subject to systematical transformational procedures to help functions of growth and developments. This lasted for a year and a half. In the beginning, middle and at the end of treatment subjects were measured so we gained absolute continuum of 747 objects (subjects) for this analysis.

We applied 14 morphological variables and it is certain they follow international biological standards, but also they are capable of covering different models of latent dimensions gained in different researches. Variables are: body height (AVIT), leg length (ADUN), arm length (ADUR), diameter of wrist (ADRZ), knee diameter (ADIK), biacromial ratio (ASIR), bilicristal ratio (ASIK), body weight (ATEZ), forearm amplitude (AOPL), lower leg amplitude (AOPK), middle amplitude of thorax (AOGK), upper arm skin folds (AKNN), back skin folds (AKNL) and stomach skin folds (AKNT).

As a hyphotesis, we do not expect significant differences between transitive points in process.

Table 1. Correlations of inferior/initial subsample

INF	AVIT	ADUN	ADUR	ADRZ	ADIK	ASIR	ASIK	ATEZ	AOPL	AOPK	AOGK	AKNN	AKNL	AKNT
AVIT	1.00	0.86	0.78	0.54	0.54	0.53	0.50	0.69	0.39	0.53	0.44	-0.28	-0.14	-0.25
ADUN	0.86	1.00	0.73	0.46	0.43	0.49	0.44	0.60	0.30	0.46	0.40	-0.25	-0.19	-0.26
ADUR	0.78	0.73	1.00	0.43	0.49	0.52	0.40	0.58	0.43	0.45	0.44	-0.24	-0.12	-0.21
ADRZ	0.54	0.46	0.43	1.00	0.61	0.49	0.42	0.61	0.47	0.46	0.45	-0.30	-0.23	-0.28
ADIK	0.54	0.43	0.49	0.61	1.00	0.43	0.39	0.65	0.55	0.55	0.54	-0.35	-0.26	-0.30
ASIR	0.53	0.49	0.52	0.49	0.43	1.00	0.48	0.59	0.44	0.48	0.49	-0.28	-0.23	-0.35
ASIK	0.50	0.44	0.40	0.42	0.39	0.48	1.00	0.60	0.38	0.44	0.46	-0.43	-0.40	-0.39
ATEZ	0.69	0.60	0.58	0.61	0.65	0.59	0.60	1.00	0.73	0.80	0.77	-0.67	-0.61	-0.70
AOPL	0.39	0.30	0.43	0.47	0.55	0.44	0.38	0.73	1.00	0.74	0.73	-0.53	-0.46	-0.54
AOPK	0.53	0.46	0.45	0.46	0.55	0.48	0.44	0.80	0.74	1.00	0.73	-0.59	-0.51	-0.59
AOGK	0.44	0.40	0.44	0.45	0.54	0.49	0.46	0.77	0.73	0.73	1.00	-0.56	-0.60	-0.67
AKNN	-0.28	-0.25	-0.24	-0.30	-0.35	-0.28	-0.43	-0.67	-0.53	-0.59	-0.56	1.00	0.78	0.76
AKNL	-0.14	-0.19	-0.12	-0.23	-0.26	-0.23	-0.40	-0.61	-0.46	-0.51	-0.60	0.78	1.00	0.80
AKNT	-0.25	-0.26	-0.21	-0.28	-0.30	-0.35	-0.39	-0.70	-0.54	-0.59	-0.67	0.76	0.80	1.00

Table 2. Correlations of middle/average subsample

MID	AVIT	ADUN	ADUR	ADRZ	ADIK	ASIR	ASIK	ATEZ	AOPL	AOPK	AOGK	AKNN	AKNL	AKNT
AVIT	1.00	0.89	0.78	0.57	0.59	0.58	0.56	0.70	0.40	0.53	0.47	-0.29	-0.19	-0.26
ADUN	0.89	1.00	0.79	0.48	0.52	0.56	0.49	0.64	0.37	0.51	0.46	-0.26	-0.19	-0.25
ADUR	0.78	0.79	1.00	0.48	0.52	0.54	0.49	0.60	0.40	0.44	0.44	-0.18	-0.11	-0.16
ADRZ	0.57	0.48	0.48	1.00	0.64	0.55	0.49	0.65	0.49	0.47	0.51	-0.34	-0.27	-0.29
ADIK	0.59	0.52	0.52	0.64	1.00	0.48	0.44	0.67	0.53	0.54	0.53	-0.37	-0.29	-0.29
ASIR	0.58	0.56	0.54	0.55	0.48	1.00	0.52	0.66	0.50	0.55	0.58	-0.28	-0.31	-0.37
ASIK	0.56	0.49	0.49	0.49	0.44	0.52	1.00	0.69	0.45	0.55	0.60	-0.47	-0.46	-0.53
ATEZ	0.70	0.64	0.60	0.65	0.67	0.66	0.69	1.00	0.75	0.80	0.82	-0.65	-0.61	-0.69
AOPL	0.40	0.37	0.40	0.49	0.53	0.50	0.45	0.75	1.00	0.74	0.76	-0.49	-0.43	-0.48
AOPK	0.53	0.51	0.44	0.47	0.54	0.55	0.55	0.80	0.74	1.00	0.75	-0.55	-0.54	-0.60
AOGK	0.47	0.46	0.44	0.51	0.53	0.58	0.60	0.82	0.76	0.75	1.00	-0.54	-0.61	-0.63
AKNN	-0.29	-0.26	-0.18	-0.34	-0.37	-0.28	-0.47	-0.65	-0.49	-0.55	-0.54	1.00	0.76	0.79
AKNL	-0.19	-0.19	-0.11	-0.27	-0.29	-0.31	-0.46	-0.61	-0.43	-0.54	-0.61	0.76	1.00	0.86
AKNT	-0.26	-0.25	-0.16	-0.29	-0.29	-0.37	-0.53	-0.69	-0.48	-0.60	-0.63	0.79	0.86	1.00

Table 3. Correlations of superior/final subsample

SUP	AVIT	ADUN	ADUR	ADRZ	ADIK	ASIR	ASIK	ATEZ	AOPL	AOPK	AOGK	AKNN	AKNL	AKNT
AVIT	1.00	0.80	0.74	0.52	0.55	0.53	0.54	0.66	0.40	0.48	0.45	-0.27	-0.18	-0.26
ADUN	0.80	1.00	0.75	0.42	0.49	0.49	0.45	0.57	0.36	0.45	0.40	-0.21	-0.15	-0.20
ADUR	0.74	0.75	1.00	0.45	0.50	0.47	0.48	0.58	0.40	0.41	0.40	-0.17	-0.09	-0.15
ADRZ	0.52	0.42	0.45	1.00	0.59	0.51	0.44	0.61	0.48	0.45	0.49	-0.34	-0.25	-0.28
ADIK	0.55	0.49	0.50	0.59	1.00	0.45	0.42	0.62	0.51	0.48	0.48	-0.34	-0.27	-0.26
ASIR	0.53	0.49	0.47	0.51	0.45	1.00	0.42	0.58	0.46	0.46	0.49	-0.21	-0.25	-0.32
ASIK	0.54	0.45	0.48	0.44	0.42	0.42	1.00	0.61	0.40	0.49	0.54	-0.43	-0.39	-0.49
ATEZ	0.66	0.57	0.58	0.61	0.62	0.58	0.61	1.00	0.72	0.72	0.74	-0.60	-0.54	-0.62
AOPL	0.40	0.36	0.40	0.48	0.51	0.46	0.40	0.72	1.00	0.68	0.74	-0.48	-0.41	-0.45
AOPK	0.48	0.45	0.41	0.45	0.48	0.46	0.49	0.72	0.68	1.00	0.71	-0.50	-0.51	-0.56
AOGK	0.45	0.40	0.40	0.49	0.48	0.49	0.54	0.74	0.74	0.71	1.00	-0.49	-0.56	-0.57
AKNN	-0.27	-0.21	-0.17	-0.34	-0.34	-0.21	-0.43	-0.60	-0.48	-0.50	-0.49	1.00	0.73	0.77
AKNL	-0.18	-0.15	-0.09	-0.25	-0.27	-0.25	-0.39	-0.54	-0.41	-0.51	-0.56	0.73	1.00	0.83
AKNT	-0.26	-0.20	-0.15	-0.28	-0.26	-0.32	-0.49	-0.62	-0.45	-0.56	-0.57	0.77	0.83	1.00

Table 4. Average correlations generated by subsample correlations

AVG	AVIT	ADUN	ADUR	ADRZ	ADIK	ASIR	ASIK	ATEZ	AOPL	AOPK	AOGK	AKNN	AKNL	AKNT
AVIT	1.00	0.85	0.77	0.54	0.56	0.55	0.53	0.68	0.40	0.51	0.46	-0.28	-0.17	-0.26
ADUN	0.85	1.00	0.76	0.45	0.48	0.51	0.46	0.60	0.34	0.47	0.42	-0.24	-0.18	-0.24
ADUR	0.77	0.76	1.00	0.45	0.50	0.51	0.46	0.59	0.41	0.43	0.43	-0.20	-0.10	-0.17
ADRZ	0.54	0.45	0.45	1.00	0.62	0.52	0.45	0.62	0.48	0.46	0.48	-0.33	-0.25	-0.28
ADIK	0.56	0.48	0.50	0.62	1.00	0.45	0.41	0.65	0.53	0.52	0.52	-0.35	-0.28	-0.28
ASIR	0.55	0.51	0.51	0.52	0.45	1.00	0.47	0.61	0.47	0.50	0.52	-0.26	-0.27	-0.35
ASIK	0.53	0.46	0.46	0.45	0.41	0.47	1.00	0.63	0.41	0.49	0.53	-0.44	-0.42	-0.47
ATEZ	0.68	0.60	0.59	0.62	0.65	0.61	0.63	1.00	0.73	0.77	0.78	-0.64	-0.59	-0.67
AOPL	0.40	0.34	0.41	0.48	0.53	0.47	0.41	0.73	1.00	0.72	0.74	-0.50	-0.44	-0.49
AOPK	0.51	0.47	0.43	0.46	0.52	0.50	0.49	0.77	0.72	1.00	0.73	-0.55	-0.52	-0.58
AOGK	0.46	0.42	0.43	0.48	0.52	0.52	0.53	0.78	0.74	0.73	1.00	-0.53	-0.59	-0.62
AKNN	-0.28	-0.24	-0.20	-0.33	-0.35	-0.26	-0.44	-0.64	-0.50	-0.55	-0.53	1.00	0.76	0.77
AKNL	-0.17	-0.18	-0.10	-0.25	-0.28	-0.27	-0.42	-0.59	-0.44	-0.52	-0.59	0.76	1.00	0.83
AKNT	-0.26	-0.24	-0.17	-0.28	-0.28	-0.35	-0.47	-0.67	-0.49	-0.58	-0.62	0.77	0.83	1.00

Table 5. Statistical significance of structural differences

Group	n	df	HI²	P
Inferior/initial	249	14	5.92	>0.05
Middle/average	249	14	8.04	>0.05
Superior/final	249	14	5.28	>0.05

Discussion and conclusion

As seen in Table 1, 2 and 3, the correlations between variables are relatively stable, which is not surprising considering that it is a morphological dimensions that in the selected sample should not significantly structurally altered. This was confirmed by testing significance (Table 5), showing that neither of the situations defined statistical significance at p = 0.05 or less.

These results show that the algorithm is stable and prepared insensitive to minor structural changes, and it is surely recommended for use in all of these situations when we expect that changes in the structure of the parameters that define the stability of a system or multivariate described by the transformation are not expecting any dramatic changes.

References

Bonacin, D. (2006). Iterim - solution for detection and optimization of variance in applied variable system. *Anthropological status and physical activity of children and youth. Interdisciplinary scientific conference. Novi Sad*, Proceedings, (pp. 289-294).

Bonacin, D. (2010). *Uvod u kvantitativne metode* [Introduction to quantitative methods. In Croatian.]. Travnik: Edukacijski fakultet.

Bonacin, D., & Bonacin, Da. (2012). Example behavior of algorithm and software for quantitative and structural multivariate differentiation of group of entities with different motor status. *Sport Science, 5*(2), 57-61.

Bonacin, D., Bilić, Ž., & Bonacin, Da. (2008). *Uvod u antropološku analizu* [Introduction to anthropological analysis. In Croatian.]. Travnik: Edukacijski fakultet.

Bonacin, D., Bilić, Ž., & Bonacin, Da. (2012). *Uvod u kineziološku analizu* [Introduction to kinesiological analysis. In Croatian.]. Travnik: Univerzitet u Travniku.

Bonacin, D., & Smajlović, N. (2005). Univerzalni model selekcije za vrhunsko sportsko stvaralaštvo [Universal model for top-level sport achievement. In Croatian.]. *Homo Sporticus, 8*(1), 36-41.

Cooley, W.W., & Lohnes, P.R. (1971). *Multivariate data analysis*. New York: John Wiley and sons. inc.

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

Johnson, A.R., & Wichern, W.D. (1992). *Applied Multivariate Statistical Anaysis*. 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* [Methods, algorithms and programs for analysis of quantitative and qualitative changes. In Croatian.]. Zagreb: Institut za kineziologiju.

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

PRIMJER PONAŠANJA ALGORITMA I PROGRAMA ZA KVANTITATIVNO I STRUKTURALNO MULTIVARIJANTNO RAZLIKOVANJE GRUPA ENTITETA RAZLIČITOG MORFOLOŠKOG STATUSA

Sažetak

Svrha ovog rada bila je priprema i provjera algoritma priređenog za analizu strukturalnih multivarijantnih razlikovanja skupina entiteta dobivenih i raspoređenih iz ukupnog efektiva. Temeljni metodološki obrazac očituje se u činjenici da unutar bilo kako definiranog uzorka postoje subuzorci koje je temeljem nekog objektivnog kriterija moguće razvrstati najmanje kao inferiorne, prosječne i superiorne, ili kod transformacijskog procesa kao početno, tranzitivno i finalno stanje. Algoritam je testiran na više primjera od kojih su zadržana dva. Jedan je prikazan u ovom radu kao primjer stabilnih subuzoraka u morfološkoj domeni koji ne pokazuju razlike u strukturi kroz transformacijski proces. Drugi primjer je objavljen u časopisu „Sport Science" u motoričkoj domeni i pokazuje upravo drastične strukturalne promjene i kako se one, upravo ovim algoritmom mogu otkriti (Bonacin & Bonacin, 2012).

Ključne riječi*: algoritam, grupe entiteta, morfologija, struktura, razlikovanje*

Assist.Prof.Dobromir Bonacin, PhD
University of Travnik
Faculty of Kinesiology
Aleja Konzula 5, 72270 Travnik, BiH
Phone: +387(0)30 517 824
E-mail: dobromirb@actakin.com