SIMPLE CREDIBILITY ANALYSIS OF MULTIVARIATE SELECTION

Dobromir Bonacin
Faculty of Kinesiology, University of Travnik, Bosnia and Herzegovina

Original scientific paper

Abstract
Based on a sample of 249 examinees, children age of 7, who were described with yet 26 variables in three occasions, we applied analysis of multivariate selection credibility procedure regarding to individual span of variables. Specially created algorithm was prepared and it indicated that the number of containing objects in such multivariate model is not and can not be justified in form of eliminatory span where aimed object exists, because at least according to one variable, almost each object can be eliminated.

Key words: *selection, span*

Introduction

Selection as a concept, to date, became standard part of procedure tools in many scientific branches (kinesiology, medicine, psychology, education, urbanism,...). In simple terms, any aimed selection of some objects from available stack, definitely is selection, no matter which purposes it was conducted for. It seems that these procedures are easiest to understand when conducted through sport; however, it is also true for medicine testing, environmental planning or professional orientation. The problem occurs when it is affirmed that some solution requires observation from different points of view, because the final object selection is influenced by more and a lot of factors that are somehow related, what usually is the case. Then, naturally, we use different existing methods for multi - criteria selection. However, selection, as a term, is actually elimination. Under assumption it is possible to determine unique set of criteria for such procedure, evidently according to that criteria, some objects will be rejected and other accepted.

The actual question is: How certain we are that the selection was correctly and methodologically conducted!? One excellent indicator there is excuse for doubting in methodological basis of selection, is exactly the way how we obtain results. Let's assume we have large number of objects presented with numerous parameters which are in various relations and some variables are metrically standardized. Then we can easily have situation that procedure leads us to really unexpected results.

So that e.g. according to norms of multivariate normal distribution, that is usually foundation of such techniques, there is real danger to make big mistakes. Algorithm was written and tested and it shows how much is this strategy justified.

Methods and 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 ($m_j= \Sigma\ xi\ /\ n$) and vector $S= (s_i\ ;\ j=1,..m)$ contains standard variable variations from V presented in X ($s_j = sqrt(\Sigma\ (xi - X)2)/n$). 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 ($z_{ij} = (x_{i,j} - m_j)/s_j$).

Under assumption we are interested in a set of objects that does not give results per any variable in some pre-determined span, we will get data copied from general space to specific space determined by parameters outside of that span. So, lets say we determined two parameters, Δ1 (lower limit of rejection) and Δ2 (upper limit of rejection) and they are marked as above mentioned Z value for easier identification.

Inside of this limits (Δ1 and Δ2) according to Gauss curve there will be certain number of data, which means that hypothetically according to each variable will be kept certain object volume. However, variations of data and interaction of variables produce model according to that amount drastically drops with only small increasing of let's say middle span, because small chances are that the same object will, according to different variables, positions inside of defined elimination span. Therefore it is interesting to study the behavior of this model in reality and how many entities we have left, depending on definition of certain elimination span. 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.

Out of 26 variable 14 were morphological 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), bicristal 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). We also applied 12 motor variables also imagined to cover area of primary motor dimensions (coordination, movement frequency, flexibility, balance, repetitive strength, explosiveness, static strength and endurance) according to different research. : side steps (MKUS), polygon backwards (MPOL), hand tapping (MTAP), foot tapping (MTAN), straddle forward band (MPRR), standing on the bench balancing (MP2O), broad jump (MSDM), throwing ball into a distance (MBLD), 20 m run from a standing start (M20V), sit-ups (MDTS), held part in the hang (MVIS), 3- min run (MT3M). After object elimination per each variable, certain set of objects remained and they are presented for two typical cases in Tables and Graphs.

Results

First situation, according to indicators in Table 1. points out elimination of average objects and it is visible that with less ration in the middle we have slow growth of remained objects and later this number is increasing exponentially (Graph 1.) It is obvious that when there is small ratio (0.22 i.e. from -0.11 do +0.11) around approx 10 % of object remains in model and more then 30% only when this ratio is totally minimized (0.10 i.e. from -0.05 do +0.05). Second situation, however, according to indicators in Table 2. point out that there is elimination of the weakest objects in defined space. It is obvious that the number of objects in beginning slowly drops, and when value on the right limit achieved -1.00, only 30 % of objects remained and on the value of -0.70 only 10 %. This is presented on Graph 2.

Discussion

In both cases obtained result clearly suggest that in multivariate selection is almost impossible to choose object if it is a large number of objects and parameters and under assumption that these parameters do not define narrow hyper cone where objects exist (because then these variables are almost the same). Realistically each object, at least according to one variable enters the chosen span and becomes eliminated. In such a way it is impossible to select objects that predict high result correspondence with goals of conducted selection. Both situations indicate that for selection procedures it is necessary to define totally new models that can endure strict criteria in a way to pick the ones that present large guarantee for positive reply to satisfy plan of selection. Described appearance because of the same limit that appears in a sense of multivariate normal distribution is equally reflected on all known models of criteria multivariate selection, standard factor, regression, canonic and other analysis which weakens this procedure and its appliance in operational purposes. It is obvious we need a step forward and we need to create models under totally new methodological conditions, so we could overcome mentioned obstacles.

Conclusion

For this research we use data of 249 examinees, children age 7, described with yet 26 variables in three occasions.

Analysis of multivariate selection credibility procedure was applied regarding to individual span of variables. Specially created algorithm was prepared and it indicated that the number of containing objects in such multivariate model is not and can not be justified in form of eliminatory span where aimed object exists, because at least according to one variable, almost each object can be eliminated It was suggested to form new selection procedures that would ensure much bigger credibility.

References

Bonacin, D. (2004). *Uvod u kvantitativne metode*. Kaštela: Vlastito.

Bonacin, D., & Smajlović, N. (2005). Univerzalni model selekcije za vrhunsko sportsko stvaralaštvo. *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 Analysis*. Englewood Cliffs: Prentice-Hall.

Malacko, J., & Rađo, I. (2004). *Tehnologija sporta i sportskog treninga*. Sarajevo: Fakultet sporta i tjelesnog odgoja.

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

Table 1.

N	L	D	O	%
1	-0.40	0.40	0	0.00
2	-0.35	0.35	1	0.13
3	-0.30	0.30	2	0.27
4	-0.28	0.28	2	0.27
5	-0.26	0.26	3	0.40
6	-0.24	0.24	3	0.40
7	-0.22	0.22	4	0.54
8	-0.21	0.21	8	1.07
9	-0.20	0.20	13	1.74
10	-0.19	0.19	15	2.01
11	-0.18	0.18	22	2.95
12	-0.17	0.17	31	4.15
13	-0.16	0.16	34	4.55
14	-0.15	0.15	36	4.82
15	-0.14	0.14	51	6.83
16	-0.13	0.13	54	7.23
17	-0.12	0.12	72	9.64
18	-0.11	0.11	89	11.91
19	-0.10	0.10	115	15.39
20	-0.09	0.09	134	17.94
21	-0.08	0.08	150	20.08
22	-0.07	0.07	166	22.22
23	-0.06	0.06	210	28.11
24	-0.05	0.05	286	38.29
25	-0.04	0.04	332	44.44
26	-0.03	0.03	409	54.75
27	-0.02	0.02	491	65.73
28	-0.01	0.01	596	79.79
29	0.00	0.00	747	100.00

Table 2.

N	L	D	O	%
1	-3.00	-3.00	747	100.00
2	-3.00	-2.75	730	97.72
3	-3.00	-2.50	710	95.05
4	-3.00	-2.25	683	91.43
5	-3.00	-2.00	653	87.42
6	-3.00	-1.75	575	76.97
7	-3.00	-1.50	493	66.00
8	-3.00	-1.25	368	49.26
9	-3.00	-1.20	334	44.71
10	-3.00	-1.15	311	41.63
11	-3.00	-1.10	274	36.68
12	-3.00	-1.05	233	31.19
13	-3.00	-1.00	199	26.64
14	-3.00	-0.95	176	23.56
15	-3.00	-0.90	160	21.42
16	-3.00	-0.85	135	18.07
17	-3.00	-0.80	102	13.65
18	-3.00	-0.75	88	11.78
19	-3.00	-0.70	71	9.50
20	-3.00	-0.65	47	6.29
21	-3.00	-0.60	41	5.49
22	-3.00	-0.55	34	4.55
23	-3.00	-0.50	31	4.15
24	-3.00	-0.45	24	3.21
25	-3.00	-0.40	15	2.01
26	-3.00	-0.35	13	1.74
27	-3.00	-0.30	7	0.94
28	-3.00	-0.25	6	0.80
29	-3.00	-0.20	6	0.80
30	-3.00	-0.15	5	0.67
31	-3.00	-0.10	4	0.54
32	-3.00	-0.05	1	0.13
33	-3.00	0.00	0	0.00

L = left border D = right border O = objects

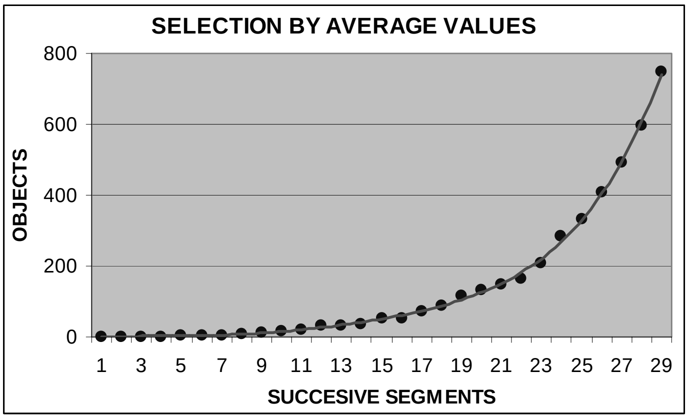

Figure 1.

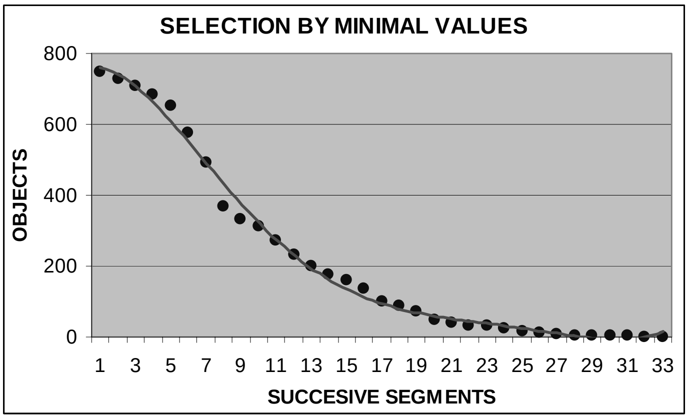

Figure 2.

JEDNOSTAVNA ANALIZA VJERODOSTOJNOSTI MULTIVARIJANTNE SELEKCIJE

Sažetak
S velikim uzorkom od 249 dječaka uzrasta 7 godina, koji su u tri navrata bili opisani sa čak 26 morfološko-motoričkih varijabli, proveden je postupak analize vjerodostojnosti multivarijantne selekcije u odnosu na individualne raspone varijabli. Pripremljen je posebno kreirani algoritam koji je pokazao da broj zadržanih objekata u takvom multivarijantnom modelu nije i ne može biti opravdan pod vidom eliminacijskog raspona u kojemu ciljani objekti egzistiraju, jer po barem jednoj varijabli gotovo svaki objekt može biti eliminiran. Predloženo je formiranje novih selekcijskih postupaka koji bi pokazivali znatno veću vjerodostojnost.

***Ključne riječi**: selekcija, rasponi*

Received: March, 12. 2008.
Accepted: May, 26. 2008.
Correspondence to:
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