ITERIM – SOLUTION FOR DETECTION AND OPTIMISATION OF VARIANCE IN APPLIED VARIABLE SYSTEM

Dobromir Bonacin PhD
Faculty of sport and physical education – University of Sarajevo

Summary

It was prepared, tested and implemented a procedure for data analysis that precisely determine portion of all variance types in applied variable system. Procedure is performed by iterative image analysis step data transformation method with reliability variance maximization and error variance minimization. This procedure quickly converges to optimal solution, and orthoblique position shows complete preservation of real parsimony structure. Behavior of algorithm was tested on more then 60 different situations and for this article purposes it was chosen a complex variant with 249 male entities first formers aged 7 described with 24 bio-motor variables.

Key words: variance, parsimony, iterations, image

Introduction

Guttmann's Theory of representation, basis for his Image theory, represents an important alternative to classical model of data analysis because it determines common part of variance in stretched space. It is also an alternative to classical measurement model, because it allows non-zero relations between error components of some composite result. On those logical basics we can define "real" result portion of any variable. Of course, we are talking about mappings our initial data to space with range of $-x_1\sigma$ do $+x_2\sigma$ ($x_{1,\ 2}$>3.5 respectively). So, this article offers an algorithm Iterim, because definition of image variables in primary range of registration is easy to derive from momentums of initial variables. On that data, repeated application of image analysis constructs more and more authenticity results with unique portion of variance getting smaller and smaller as iterations continued. Model obviously converges to total common variability without composing linearly dependent variables, with all basic relations remained.

Methods

<u>Algorithm ITERIM</u>

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 results of some test T and 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

$$\mathbf{Z} = \mathrm{E} \otimes \mathrm{P} \quad \Big| \quad \mathbf{Z}^{\mathbf{t}}\ \mathrm{e} = 0\ , \quad \mathrm{dg}\ (\mathbf{R}) = \mathbf{I}$$

where e is summation vector (n,1), and $\mathbf{R} = \mathbf{Z}^{\mathbf{t}}\ \mathbf{Z}$ is correlation matrix of results in test T, and **Z** is matrix of standardized results in variables from V. Let **H** is some unknown matrix of real results of those variables and let **U** is also unknown matrix of error variables concatenated to variables from V. Basic model of measurement is

$$\mathbf{Z} = \mathbf{H} + \mathbf{U}.$$

Guttmann's evaluation of maximal variance, unique portion of each variable from V will be diagonal elements of unique matrix from V i.e. $\mathbf{U}^2 = (\mathrm{dg}\ \mathbf{R}^{-1})^{-1}$, because really elements from $\mathbf{U}^2$ represents vectors orthogonal on corresponding vectors from V existing in common space. Owing to that, results that are consequence of measurement errors is possible to solve with simple operation

$$\mathbf{E} = \mathbf{Z}\ \mathbf{R}^{-1}\ \mathbf{U}^2$$

Those are projections of results on set of vectors that represent error variance of measurement. Matrix of real results is

$$\mathbf{H} = \mathbf{Z} - \mathbf{E} = \mathbf{Z}\,(\,\mathbf{I} - \mathbf{R}^{-1}\ \mathbf{U}^2\,).$$

Variable from H are ***image*** variable, and variable from U are ***anti-image*** variable. But, when we dispose with composite instruments, image variables matches real variables, and anti-image variables matches variables of measuring errors. Furthermore, dg $(\mathbf{G}) = \mathbf{I} - \mathbf{U}^2$, so variance of image variables are just variances of real test results, and exactly determinacy coefficients of each variable estimated with other variables from V. This certainly worth with set of objects described with some casual variables. Their common variance in image model ensures maximal dose of common variability in stretched space. Guttmann's image variables are

$$\mathbf{H} = \mathbf{Z} - \mathbf{E} = \mathbf{Z}\,(\,\mathbf{I} - \mathbf{R}^{-1}\ \mathbf{U}^2\,).$$

Matrix of covariance is G, with determinacy coefficients in main diagonal. Spectral decomposition of that matrix derive eigen vectors X. Operation W = Z * X with norm condition (**M,** σ)**,** where M is a mean and σ is a standard deviation, determines image variables in basic variables metric exactly in a part that is common to such variables set. Obviously, variables in W are defined with common portion of basic metric from set in G.

Repeated application of image analysis on W estimate better and better representative common variability. Procedure converges to maximum (i.e. 1.00). Criterion for process ending is defined in iteration when

$$h^2 >= 0.95 \quad \text{and} \quad r^2 >= 0.99,$$

when communality (h^2) of each particular variable is equal or bigger then 0.95, and reliability (r^2) (communality + specificity) of whole set is equal or bigger then 0.99.

Numerical example

The sample was comprised of 249 children, primary school first formers from Split, who, at the beginning of the experimental procedure, were 7 years +/- 2 months old. All the children had no visible aberrations, and they were all able to participate in a normal program of work in primary school. **The sample of variables** necessary for the assessment was selected in such a way as to cover both the morphological and the motor status: body height (AVIT), arm length (ADUR), leg length (ADUN), biacromial width (ASIR), knee diameter (ADIK), wrist diameter (ADRZ), body weight (ATEZ), chest circumference (AOGK), lower leg circumference (AOPK), forearm circumference (AOPL), skinfold of the back (AKNL), upper arm skinfold (AKNN) and abdominal skinfold (AKNT). All the measures were taken according to the international biological program. The following variables were used for the assessment of the motor status: side steps (MKUS), held part in the hang (MVIS), long jump from a standstill (MSDM), standing on the bench (MP20), polygon backwards (MPOL), sit-ups (MDTS), 20m run from a standing start (M20V), straddle forward bend (MPRR), hand-tapping (MTAP), foot-tapping (MTAN) and throwing the ball for distance (MBLD), 3-min run (FT3M) was used to assess the aerobic work. All the measurements were done by qualified people who had significant experience in collecting the aforementioned initial data. For common variability maximisation it was Iterim procedure applied. Stabilization occurs (as in other cases) after maximally 10 iterations.

Results

Inspection of Table 1 shows that "real" parts of results become evidently higher with its portion significantly bigger (from 67.88 to 96.70 %). Error component was reduced from 11 % to 0.1 %, with only 3.2 % of specificity of unique part of measuring instrument variance. These data shows that Iterim procedure maximizes total part of common variability in system. Initial manifest results are cleared from additives characteristic with systematic and un-systematic errors.

Of course, those results components from Table 1 do not have to be enough persuasive proof of Iterim model credibility for everyone. For that reason, data were processed with orthoblique solution before and after Iterim application for model evaluation. As we can see (table 2.), all maximal projections are perfectly preserved after iterations, and maximal variable projections for OBQ factors identification define all latent dimensions recognized as: longitudinal bone growth (AVIT, ADUN, ADUR), transversal bone growth (ADRZ, ADIK, ASIR, ATEZ), volume growth (AOPL, AOPK, AOGK), adipose tissue (AKNN, AKNL, AKNT), motor information output (MKUS, MPOL, MP2O, MTAP, MTAN, MSDM) and motor energetic output (MSDM, MBLD, M20V, MDTS, MVIS, MT3M). Correlations of OBQ factors are logically identical, too.

FIRST ITERATION					
VAR.	h2	u2	s2	e2	r2
AVIT	0.537	0.463	0.249	0.215	0.785
ADUN	0.536	0.464	0.249	0.215	0.785
ADUR	0.458	0.542	0.248	0.294	0.706
ADRZ	0.763	0.237	0.181	0.056	0.944
ADIK	0.559	0.441	0.247	0.195	0.806
ASIR	0.528	0.472	0.249	0.223	0.777
ATEZ	0.647	0.353	0.228	0.125	0.875
AOPL	0.544	0.456	0.248	0.208	0.793
AOPK	0.494	0.506	0.250	0.256	0.744
AOGK	0.557	0.443	0.247	0.196	0.804
AKNN	0.721	0.279	0.201	0.078	0.922
AKNL	0.725	0.275	0.200	0.076	0.924
AKNT	0.605	0.395	0.239	0.156	0.844
MKUS	0.772	0.228	0.176	0.052	0.948
MPOL	0.715	0.285	0.204	0.081	0.919
MP2O	0.738	0.262	0.193	0.069	0.931
MTAP	0.807	0.193	0.156	0.037	0.963
MTAN	0.814	0.186	0.151	0.035	0.966
MSDM	0.751	0.249	0.187	0.062	0.938
MBLD	0.722	0.278	0.201	0.077	0.923
M20V	0.742	0.258	0.192	0.067	0.933
MDTS	0.863	0.137	0.118	0.019	0.981
MVIS	0.833	0.167	0.139	0.028	0.972
MT3M	0.860	0.140	0.121	0.020	0.980
SUM	16.292	7.708	4.872	2.836	21.164
PART%	67.884	32.116	20.299	11.817	88.184

TENTH ITERATION					
VAR.	h2	u2	s2	e2	r2
AVIT	0.958	0.042	0.040	0.002	0.998
ADUN	0.969	0.031	0.030	0.001	0.999
ADUR	0.955	0.045	0.043	0.002	0.998
ADRZ	0.962	0.038	0.037	0.001	0.999
ADIK	0.953	0.048	0.045	0.002	0.998
ASIR	0.958	0.042	0.040	0.002	0.998
ATEZ	0.963	0.037	0.036	0.001	0.999
AOPL	0.963	0.037	0.036	0.001	0.999
AOPK	0.957	0.043	0.041	0.002	0.998
AOGK	0.960	0.040	0.038	0.002	0.998
AKNN	0.968	0.032	0.031	0.001	0.999
AKNL	0.975	0.025	0.025	0.001	0.999
AKNT	0.962	0.039	0.037	0.002	0.999
MKUS	0.978	0.023	0.022	0.001	1.000
MPOL	0.974	0.026	0.025	0.001	0.999
MP2O	0.964	0.036	0.035	0.001	0.999
MTAP	0.971	0.029	0.028	0.001	0.999
MTAN	0.981	0.019	0.019	0.000	1.000
MSDM	0.969	0.031	0.030	0.001	0.999
MBLD	0.958	0.042	0.041	0.002	0.998
M20V	0.969	0.031	0.030	0.001	0.999
MDTS	0.980	0.020	0.019	0.000	1.000
MVIS	0.976	0.025	0.024	0.001	0.999
MT3M	0.987	0.013	0.013	0.000	1.000
SUM	23.208	0.792	0.764	0.028	23.972
PART%	96.701	3.299	3.183	0.117	99.883

Table 1: Results components
(h^2 = communality, u^2 = unique, s^2 = specificity, e^2 = error, r^2 = reliability)

PRVA ITERACIJA						
	OBQ1	**OBQ2**	**OBQ3**	**OBQ4**	**OBQ5**	**OBQ6**
AVIT	-0.05	0.17	-0.14	-0.01	-0.07	**0.59**
ADUN	-0.11	-0.11	0.06	0.02	0.18	**0.64**
ADUR	0.32	-0.07	0.02	0.10	-0.24	**0.57**
ADRZ	-0.19	**0.82**	0.08	-0.25	0.13	-0.09
ADIK	-0.07	**0.60**	-0.06	0.09	0.02	0.07
ASIR	0.08	**0.35**	-0.02	-0.07	0.03	**0.41**
ATEZ	-0.01	**0.63**	0.19	0.23	-0.05	0.09
AOPL	-0.15	0.10	-0.07	**0.62**	0.13	-0.09
AOPK	0.16	0.09	-0.10	**0.58**	-0.20	0.08
AOGK	-0.07	-0.09	0.21	**0.48**	0.39	0.03
AKNN	0.00	0.14	**0.77**	-0.04	-0.01	0.02
AKNL	0.02	0.01	**0.81**	0.05	0.04	-0.07
AKNT	0.08	-0.06	**0.68**	-0.01	-0.13	-0.01
MKUS	**-0.67**	0.05	0.09	0.18	-0.08	-0.17
MPOL	**-0.49**	-0.28	0.21	0.23	-0.04	0.08
MP2O	**0.79**	0.25	-0.20	0.28	-0.31	-0.24
MTAP	**0.83**	-0.13	0.15	0.04	0.11	0.06
MTAN	**0.87**	-0.10	0.12	0.04	0.04	0.01
MSDM	**0.40**	0.14	0.02	-0.20	**0.28**	-0.12
MBLD	0.07	0.13	-0.01	0.03	**0.72**	-0.02
M20V	-0.27	0.10	-0.16	0.09	**-0.63**	-0.07
MDTS	0.28	0.15	-0.07	0.08	**0.61**	-0.16
MVIS	-0.12	-0.11	-0.15	0.16	**0.89**	-0.03
MT3M	-0.15	-0.03	-0.03	-0.08	**0.92**	0.13
	OBQ1	**OBQ2**	**OBQ3**	**OBQ4**	**OBQ5**	**OBQ6**
OBQ1	1.00	0.15	-0.27	-0.09	0.64	-0.07
OBQ2		1.00	0.00	0.13	0.28	0.45
OBQ3			1.00	0.11	-0.40	-0.03
OBQ4				1.00	-0.01	-0.10
OBQ5					1.00	-0.10
OBQ6						1.00

DESETA ITERACIJA						
	OBQ1	**OBQ2**	**OBQ3**	**OBQ4**	**OBQ5**	**OBQ6**
AVIT	-0.09	0.23	-0.20	-0.03	-0.08	**0.80**
ADUN	-0.16	-0.13	0.06	-0.02	0.24	**0.88**
ADUR	0.44	-0.18	0.04	0.14	-0.32	**0.88**
ADRZ	-0.19	**0.97**	0.09	-0.32	0.13	-0.11
ADIK	-0.06	**0.77**	-0.09	0.16	0.02	0.13
ASIR	0.09	**0.43**	0.01	-0.08	0.11	**0.58**
ATEZ	0.00	**0.76**	0.26	0.27	-0.07	0.12
AOPL	-0.17	0.10	-0.09	**0.86**	0.10	-0.11
AOPK	0.20	0.12	-0.12	**0.85**	-0.26	0.04
AOGK	-0.11	-0.17	0.32	**0.68**	0.53	0.03
AKNN	0.01	0.23	**0.85**	-0.03	-0.05	-0.01
AKNL	-0.03	0.00	**0.97**	0.03	0.11	-0.07
AKNT	0.07	-0.11	**0.88**	-0.04	-0.13	0.02
MKUS	**-0.76**	0.03	0.12	0.19	-0.08	-0.19
MPOL	**-0.53**	-0.39	0.26	0.20	-0.09	0.22
MP2O	**0.86**	0.25	-0.17	0.27	-0.29	-0.27
MTAP	**0.86**	-0.13	0.15	0.03	0.16	0.03
MTAN	**0.94**	-0.13	0.14	0.02	0.07	0.02
MSDM	**0.37**	0.08	0.07	-0.25	**0.46**	-0.11
MBLD	0.07	0.16	-0.03	0.04	**0.83**	-0.03
M20V	-0.29	0.11	-0.15	0.08	**-0.72**	-0.12
MDTS	0.30	0.16	-0.09	0.10	**0.63**	-0.19
MVIS	-0.14	-0.13	-0.22	0.20	**0.91**	-0.02
MT3M	-0.24	-0.04	-0.06	-0.12	**1.00**	0.16
	OBQ1	**OBQ2**	**OBQ3**	**OBQ4**	**OBQ5**	**OBQ6**
OBQ1	1.00	0.13	-0.25	-0.05	0.60	-0.04
OBQ2		1.00	-0.03	0.17	0.26	0.45
OBQ3			1.00	0.07	-0.36	-0.04
OBQ4				1.00	0.03	-0.11
OBQ5					1.00	-0.09
OBQ6						1.00

Table 2: Image-orthoblique factors before and after ITERIM procedure
(OBQ1-6 = skew rotated assembly with factor correlations)

Discussion

Iterim model shows that with relatively small time and resources consumption, data measured in real metric can be easily and simple iteratively re-parameterized on image metrics with extremely clean set of data. Generally, common part of variability, of course, depends of applied variables, but once when space is stretched and variables determined, this procedure for surely maximizes everything that in such space exists. Minimization of errors in results is absolutely ensured.

Everything mentioned, allow that we can namely proclaim Iterim as best procedure for initial formal data application, especially for preparation of multivariate procedures that intend to explicate any result and generate any type of higher level indicators. That is very important for objective interpretation of results and application of comprehensions in operational sense.

Literature

1. Bala, G. (1990). Logičke osnove metoda za analizu podataka iz istraživanja u fizičkoj kulturi. Sopstveno izdanje. Novi Sad.
2. Bonacin, D. (2004). Uvod u kvantitativne metode. Vlastito izdanje, Kaštela.
3. Johnson, R. A., Wichern, D. W. (1992). Applied multivariate statistical analysis. Prentice-Hall.
4. Rađo, I., Wolf, B. (1998). Analiza grupisanja manifestnih varijabli. FFK, Sarajevo.

Sažetak

Pripremljen je, testiran i implementiran postupak analize podataka koji precizno determinira udio svih vrsta varijanci u sustavu odabranih varijabli. Postupak je izveden na temelju iterativne metode image analize pri čemu se inicijalni podaci u koracima transformiraju tako da se maksimizira prava varijanca, a minimizira error i specificitet. Dokazano je da postupak asimptotski brzo konvergira potpunom rješenju, te da su orthoblique faktorski obrasci u potpunosti očuvani uz maksimizaciju parsimonijske strukture. Ponašanje algoritma je testirano na više od 60 različitih situacija, a za ovaj primjer odabrana je vrlo složena varijanta koja sadrži 249 muških entiteta uzrasta 7 godina opisanih sa 24 biomotoričke varijable na samom početku osnovne škole.

Ključne riječi: varijanca, parsimonija, iteracije, image