TOOLS FOR THE ANALYSIS OF THE STRUCTURE AND DYNAMICS OF ATHLETIC DEVELOPMENT ON THE EXAMPLE OF HANDBALL

by

Igor Rygula[1]

The results of a two-year observation[2] of female handball players were processed using multidimensional statistical analysis. In accordance with the assumptions of system analysis and praxeologic methodology, the issues of the selection and choice of girls for sports training are presented on the example of the initial and directed stage of the training process. The work demonstrates that the fundamental issue in the sport science is the proper choice of the variables and their indices that best describe the relevant phenomenon. The solution of this problem enables the effective analysis of the structure and dynamics of the phenomena the sports training. To achieve this aim, a complementary set of the statistical analysis tools was presented, such as linear sorting of the investigated objects, cluster analysis, factor analysis and discriminative analysis.

Key words: synthetic development index, discriminative analysis, object agglomeration

Introduction

The reality around us is complex and multidimensional by nature, and situations, where a single variable is sufficient to explain the relevant phenomenon, are exceptional. Therefore, in order to describe, investigate given reality, a social or biological phenomenon, we should describe it as certain representation of the set of features. The better we know a given problem, the more successful will be the specified vector of the variables describing it. Our research problem most often is to determine the relations on

[1]Prof. dr , Chair of System Analysis in Sports, University School of Physical Education, Katowice, Poland

[2]Data from the investigation in the framework of KBN grant "Optimization of sport selection"

the set of these features (among these variables). Most often they have a specific character (deterministic, probabilistic, stochastic or strategic, based on game theory), therefore for the analysis of them we should use the complementary set of measuring and analytical tools. If the set of the analyzed variables is a subset of E elements, and relations on this set will constitute the relational structure S^R, we may treat them as ordered pair of the so-called relational system M^R, therefore:

$$M^R = \langle E, S^R \rangle$$

As the above equation implies, to solve the research problem, we should first determine the structure of investigated objects and the kind of relations between them. However, it should be stressed that the aim of the scientific investigation is to obtain the knowledge of cause-effect relations, therefore it is necessary to activate a number of mechanisms preventing detection of the so-called virtual relations (Nachmias 2001). Generally we assume that from the vast chaos of phenomena we are investigating, we will be able to select the elements of importance and isolate them from less important issues and to declare with certain accuracy that given flow of phenomena is caused by specified causes (Rygula 2001, Campbell and Stanley 1963).

The main objective of this paper is to little known and rarely used in sport science tools of statistical analysis, in order to gain significant information on the investigated objects and in this way obtaining answers for the following research questions:

1. How female handball players may be linearly sorted?
2. What is the development dynamics of 13-17 year old female handball players?
3. What is the structure of talent of the investigated athletes?
4. Which variables show the greatest discriminative power in handball team?

Material, methods and tools[3] of investigation

The obtained research material, aimed at the application of the discussed statistical analysis tools, being the base for obtaining answers to the research questions formulated above, was formed by 396 female handball players (aged 13 to 17), taking part in three stages of the experimental model (Rygula and Jarzabek, 2000). In this way, to the analysis a set of data was used of five groups of athletes, three half-year measuring periods of 26 variables. The merits of the case were the base for the choice of the measurement results of the given group for statistical analysis.

To obtain the answers to the formulated research questions, the experimental method and direct observation method were used. The investigation scheme of $RX_n^n\ Y_n^n$ was used, i.e. more than one dependent variable (Y_n), *n* independent variables (X_n), respecting randomization principle (R).

Results

In order to obtain the answers to the above questions, the distribution of the analyzed features should be determined first, to find, whether the athletes of the separate age groups correctly represent the mass of the female handball players of respective skill level. The analyses of positional measures, variation and distribution were made for all groups, while this paper presents the results for group of females born in 1982 (group A) only, from stage III of the investigation. They are presented in Table 1.

The analysis of the computed variation coefficients (V) indicates that in the investigated group A (year of birth 1982), the greatest variation is observed in the feature x_{23} (index of the special coordination skills), x_6 i.e. the variable describing the drop of power in the Wingate test. The indicated variables have shown the greatest differentiation during the full period of the pedagogic experiment (i.e. their variation indices (V) obtained their greatest values in the

[3] This paper does not contain the description of measuring and analytical tools, as they were in sufficient detail presented in another paper (Rygula, Jarzabek (2000): The diagnostic value of the analytic tools in the young people handball)

I, II and III series of the investigation). The lowest differentiation was observed for the features x_2, i.e. body height and x_9, i.e. starting speed. These variables have shown low differentiation during the whole period of the experiment.

Assuming after Dziembala (1975) that the feature has Gaussian distribution, when the asymmetry index A_s ∈ <–2, 2>, we may say that most of the variables used in the investigation have a Gaussian distribution. Only the x_2 variable, i.e. the body height shows temperate asymmetric left-sided distribution. Assuming that the kurtosis factor K_u ∈ <–3, 3>, we may find the difficulty of the chosen tests for the investigated athletes. The appropriate analysis of the distribution of the investigated variables (A_s and K_u) convinces that the girls taking part in the investigation represent well the population of 13-year old handball players, and to the analysis of the evaluated features the so-called strong statistical test may be used.

Table 1. The descriptive parameters of the distribution of the investigated features obtained by the female athletes in stage III (August 1997) of a two-year pedagogic experiment in group A (year of birth 1982)

No. (x_j)	**VARIABLE**	**X**	**S**	$\overline{V}$	A_s	K_u
1.	Body mass (kg)	54.894	7.956	14.5	-1.981	2.355
2.	Body height (cm)	165.394	5.949	*3.6*	-2.790	2.911
3.	Fat content (%)	20.667	3.054	14.8	0.203	1.069
4.	Maximum power (W/kg)	7.763	0.781	10.1	-0.918	1.974
5.	Total work (J/kg)	98.424	10.765	10.9	-1.457	2.411
6.	Fatigue index (%)	3.595	2.015	**43.3**	0.255	2.123
7.	Time to obtain maximum power output (s)	6.454	1.395	16.6	1.788	1.878
8.	Duration of maximum power (s)	3.874	1.656	23.5	1.413	1.962
9.	Starting speed – 5m (s)	1.205	0.076	6.3	1.030	2.074
10.	Maximum speed - 10m (s)	1.516	0.065	*4.3*	-1.189	2.434
11.	Explosive strength of lower limbs (cm)	173.848	14.462	8.3	-0.667	1.498
12.	Explosive strength of upper limbs (m)	9.203	1.044	11.3	-1.478	2.445
13.	Agility (s)	11.778	0.563	4.8	-0.556	1.323
14.	Endurance - shuttle run (stages)	7.909	1.373	17.4	0.001	1.071
15.	Suppleness (cm)	24.242	3.542	14.6	1.245	2.609
16.	PWC_{170} (W/kg)	1.924	0.342	17.8	-0.032	1.041
17.	Index of basic technique in version I - $W_{TP}1$ (pts)	30.606	3.420	11.2	0.571	1.332
18.	Index of special motor skills - $W_{SUM}1$ (s)	65.604	3.656	5.6	0.573	1.337
19.	Index of special aptitude - $W_{SS}1$ (pts)	0.470	0.075	15.9	0.320	1.120
20.	Index of basic technique in version II - $W_{TP}2$ (pts)	30.303	3.298	10.9	-0.808	1.724
21.	Index of special motor skills - W_{SUM} 2 (s)	70.163	5.297	7.5	0.864	1.756
22.	Index of special aptitude - W_{SS} 2 (pts)	0.437	0.079	18.0	-1.283	2.230
23.	Index of coordinative special aptitudes -W_{KZS} (s)	4.558	2.233	**49.0**	1.944	2.779
24.	Index of synthetic special aptitude - W_{SSS} (pts)	0.453	0.076	16.7	-0.398	1.207

25.	Intelligence (spatial imagination) (pts)	27.152	8.877	32.7	0.898	1.862
26.	Index of play effectiveness (%)	34.58	10.594	32.2	-0.396	1.802

1. Linear sorting of investigated objects (girl athletes)

Methods of linear sorting of the objects enable to set the linear hierarchy among the analyzed group of athletes. This enables to decide which girl is best, which is second best, etc., using aggregate criterion. Methods of such type seem to be especially interesting for team plays, when there is no score which could be assigned to specific players, especially placed on different positions in the team. The main issue is to define in possibly accurate way the so-called general criterion (e.g. "the level of aptitudes" of the athlete, the level of training - development, etc.) and to define a set of statistical features, measuring different aspects of this general criterion. The choice of features is of course of basic importance and different sets of features generally result in different hierarchies. Where there is a great number of diagnostic traits, we may chose the representative ones or use another approach (Hellwig 1968, 1968). Such procedure eliminates repeating (co-linearity) of similar information by different traits.

Numerous authors have suggested many tools for linear sorting of investigated objects (Grabinski 1992). When we decide for a specific measure, we should be aware of the numerous choices. The tool used requires computation of an arithmetic mean of diagnostic variables, which were made comparable through unilaterization and expressing this mean on point scale in the interval <0; 100>. The relevant expression is

where m is the number of considered features, a_j is the weight of the j-th variable and the summed values were normalized with the use of formulae presented in paper (Rygula 2000, p. 112)

We will first present the computation of the synthetic index W_i on the basis of the full list of features from Table 1. This formula requires first the determination of the character of the features (stimulant/nominant/destimulant) and then the determination of the reference points (lower and upper). The character of the features was set by the substantial analysis, while the reference point were generally assumed by slight rounding down the minimum value determined from whole statistical material (all measurements of all age groups), and by rounding up the observed maximum values. These settings are shown in Table 2.

Table 2. The character of the features and reference points used for the construction of the synthetic index Wi - index of athletic development

Investigated features	**Unit**	**Character of the feature**	**Lower reference value**	**Upper reference value**
1. Body mass	kg	S	30	85
2. Body height	cm	S	150	200
3. Fat content	%	D	10	35
4. Maximum power	w/kg	S	5	10
5. Total work	J/kg	S	60	130
6. Power drop index	%	S	0	12
7. Time to obtain maximum power	s	D	3,5	14,0
8. Duration of maximum power	s	S	1	10
9. Starting speed - 5m	s	D	1,0	1,4
10. Maximum speed - 10m	s	D	1,2	1,8
11. Explosive strength of lower limbs	cm	S	140	240
12. Explosive strength of upper limbs	cm	S	7	15
13. Agility	s	D	9	13
14. Endurance - shuttle run	stages	S	4	11
15. Suppleness	cm	S	8	34
16. PWC_{170}	W/kg	S	1,3	8,1
17. Basic technique - $W_{TP}1$	pts	S	2,0	42
18. Special motor skills - $W_{SUM}1$	s	D	50	80
19. Special aptitude - $W_{SS}1$	pts	S	0,25	0,85
20. Basic technique - $W_{TP}2$	pts	S	20	45
21. Special motor skills - W_{SUM} 2	s	D	50	90
22. Special aptitude - W_{SS} 2	pts	S	0,20	0,85
23. Coordinative special aptitudes -W_{KZS}	s	D	0	13
24. Synthetic special aptitude - W_{SSS}	pts	S	0,20	0,85
25. Intelligence (spatial imagination)	pts	S	10	60
26. Play effectiveness	%	S	15	80

(S - stimulant; D - destimulant)

Presented are computations for the group of athletes born in 1981, for three measurements.

Table 3. Synthetic indices and ranks in the group of girl athletes born in 1981

Number of the athlete	Index Measure-ment 1	Index Measure-ment 2	Index Measure-ment 3	Rank Measure-ment 1	Rank Measure-ment 2	Rank Measure-ment 3
1	49,7	54,8	55,8	12	15	16
2	46,6	48,8	51,6	19	24	22
3	40,3	48,3	50,4	31	26	24
4	46,6	51,5	52,5	18	20	21
5	40,9	49,1	54,6	30	23	19
6	39,8	41,9	40,9	32	35	36
7	44,2	46,3	45,9	24	28	28
8	55,2	61,0	61,9	3	5	5
9	46,1	46,9	48,3	21	27	26
10	46,7	56,8	60,6	16	11	7
11	46,8	54,3	56,6	15	18	14
12	44,2	54,6	57,6	23	17	13
13	52,3	57,6	57,9	9	8	12
14	36,9	44,8	45,2	35	32	29
15	39,5	44,2	45,1	33	33	30
16	43,1	45,0	43,8	26	31	31
17	41,0	43,8	43,4	29	34	34
18	51,3	57,5	61,8	10	9	6
19	48,6	56,7	55,9	13	12	15
20	45,9	50,6	53,8	22	21	20
21	55,0	58,4	58,8	4	7	11
22	44,0	55,6	58,8	25	13	10
23	41,4	45,7	43,6	28	29	33
24	56,7	64,0	62,2	1	3	4
25	36,4	38,4	39,7	36	37	37
26	35,4	39,9	41,7	37	36	35
27	46,5	49,9	51,5	20	22	23
28	50,9	58,8	55,2	11	6	17
29	53,1	55,0	58,8	8	14	9

30	53,2	61,5	63,1	7	4	3
31	47,6	54,6	54,8	14	16	18
32	55,3	65,9	64,0	2	2	2
33	53,4	66,6	70,1	6	1	1
34	41,7	48,3	48,0	27	25	27
35	37,5	45,4	43,7	34	30	32
36	46,7	52,9	49,7	17	19	25
37	53,8	57,5	60,6	5	10	8
38	34,9	37,5	37,5	38	38	38

The seemingly simple computations have yielded very interesting index, enabling to evaluate the level and dynamics both of the single athletes and the group in whole. For instance, athlete 1 has growing index, but the group is developing faster, and as an effect, her relative position drops from 12 to 15 and then 16. The athlete 12 has improved her index by over 10 points and advanced from position 23 to position 13. The leader in two last measurements is athlete 33, the only one to obtain the index of over 70.

Table 4 shows the descriptive characteristics of the indexes for whole group.

Table 4. Descriptive characteristics of the indexes of the group born in 1981

Measurement	Mean	Minimum	Lower quartile	Median	Upper quartile	Maximum	Standard deviation
1	46,0	34,9	41,0	46,5	51,3	56,7	6,2
2	51,8	37,5	45,7	52,2	57,5	66,6	7,5
3	52,8	37,5	45,2	54,2	58,8	70,1	7,9

Source: own computations

The data in the above table indicate that the progress in the first period (between measurement 1 and 2) was greater than in the second period (between measurement 2 and 3). The significant development is observed in only "upper" 70% of the athletes, which causes the increased differentiation of the group.

The observation of the changes of the synthetic index W_i provide interesting findings. The distribution moves right (which is confirmed by the characteristics of this index in Table 4), but at the same time it loses its Gaussian character, because the athletes are developing with different rate.

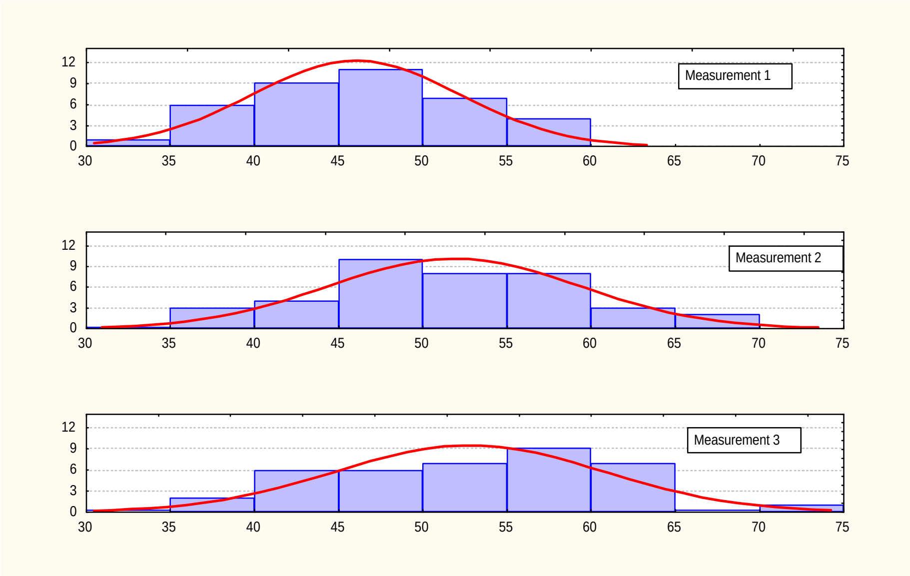

Fig. 1. The dynamics of the distribution of the synthetic index

2. Cluster analysis

The cluster analysis enables grouping of objects or grouping of features. In both cases we find the subsets of more uniform taxonomic units. In reference to the features they are the subsets of the features carrying a similar information on objects (athletes). These are most often correlated features. If we are analyzing objects, we obtain subgroups of similar (in relation of assumed features) athletes. This enables the identification of certain "types" of athletes, which in further proceedings may provide the basis for the differentiation of the kind and intensity of exercise.

We will now present an example of using cluster analysis for grouping the athletes. We will take our data from the first measurement of group B "1981", Here we have 38 objects (athletes) characterized with 26 features. As the features are measured in different units, they must be first made comparable. Here we use the classic standardization formula: *(value of the* *feature – arithmetic mean)/standard deviation.* In this way, we obtain dimensionless values, almost all contained in the (–4, +4) range.

As the original variables are continuous in character, it is justified to use the Euklidean distance as the measure of distance used in the analysis.

In accordance with the results of the simulation investigation, called upon in the description of taxonomic methods (Rygula 2001) and almost standard research practice, as an agglomeration procedure, the Ward technique has been chosen. In this way we obtain the following dendrogram.

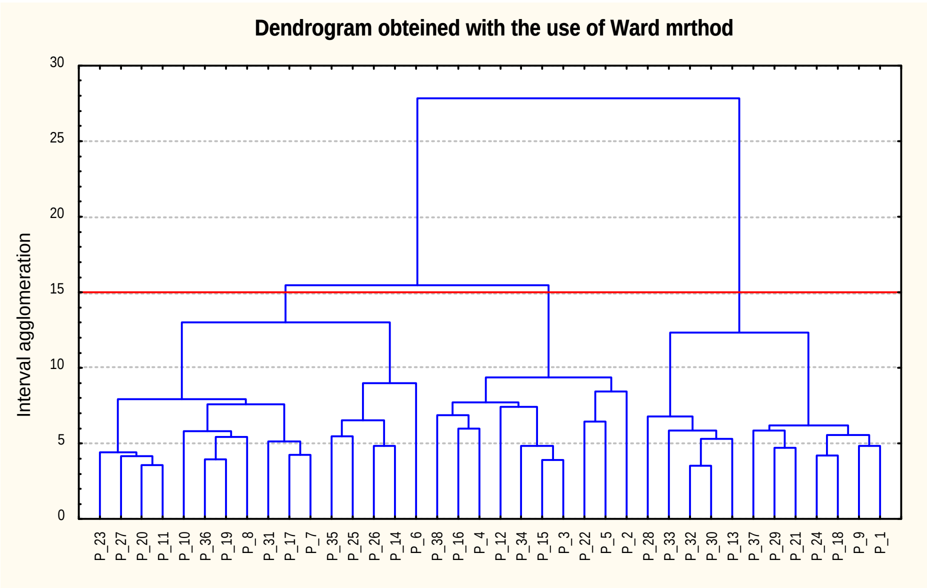

Fig. 2. Dendrogram obtained with the use of Ward method

On this figure, the separate athletes are numbered in the sequence of the table of data. The taxonomic papers describe over 35 methods of stopping the agglomeration process (dendrogram "sectioning") with the aim of finding the optimum division of the analyzed set. No single proposal has yet found general recognition and it is admissible to section the dendrogram on the basis of the evaluation of its structure. In our example, it seems reasonable to assume the agglomeration distance of 15 as the sectioning level. In this way three groups of athletes are formed. The first one (starting from the left hand side of the dendrogram) contains 11 athletes (number 23 to number 7 in the figure), second group contains 15 athletes (number 35 to number 2), while the third group contains 12 athletes (28 to 1). It is evident that the numbering is arbitrary and has no substantial meaning.

Let's have a closer look at these subgroups. The simplest method is to compute the mean value of features in the subgroups, compare them (test the equality) and then try to interpret them. The statistical test used for comparing mean values, when the number of groups is greater than two, is single-factor variance analysis. Its correct use requires the meeting of the assumption on the Gaussian distribution of investigated features and the equality of variance in the groups. When the assumption on the Gaussian distribution is not true, it is recommended to use the Kruskal-Wallis test, which is non-parametric counterpart of the variance analysis. In our considerations the results of both tests will be presented. The modern method of growing popularity for the presentation of the results of statistical tests is the use of p values (the so-called test probability). If this value is less than the assumed significance level a, the tested zero hypothesis should be rejected and the alternative hypothesis should be chosen.

Assuming the significance level of 0.10 (which seems to be justified by the relatively small sample), for the majority of the features the results of both versions of the variance analysis are the same. The differences are observable only for the *body height* and *maximum speed*. The final decision on the evaluation of the similarity of the means of these features may be taken after testing, how close the distribution approximates the Gaussian distribution. Shapiro-Wilk test was used here. For *body height*, the *p* values in this test are respectively: Group 1 - 0.6410; Group 2 - 0.0045; Group 3 - 0.4807. It is evident that in group 2 the distribution of the *height* clearly differs from the Gaussian distribution, therefore in this case the indications of the nonparametric test are more trustworthy. For the maximum speed, the following *p* values were obtained: Group 1 - 0.3574; Group 2 - 0.7309; Group 3 - 0.3082. None of them is smaller than the significance level of 0.10, therefore we may use the indications of the parametric analysis of variance (stronger test than the nonparametric counterpart). In Table 5 the rows corresponding to the means were significantly differentiated. A great number of significant differences indicate correct grouping.

Table 5. Comparison of three groups of athletes on the basis of the results of the cluster analysis

Feature	**Unit**	**Mean Group 1**	**Mean Group 2**	**Mean Group 3**	***p* ANOVA**	***p* Kruskal-Wallis**
Body mass	kg	54,7	57,1	55,9	0,5001	0,2362
Body height	cm	169,3	167,7	165,2	0,1396	0,0388
Fat content	%	19,9	21,8	19,6	0,0114	0,0174
Maximum power	w/kg	7,2	6,9	7,8	0,0104	0,0189
Total work	J/kg	90,0	86,9	97,3	0,0124	0,0072
Fatigue index	%	2,9	1,7	4,1	0,0551	0,0121
Time to obtain maximum power	s	7,9	9,2	8,0	0,1545	0,6930
Duration of maximum power	s	4,5	4,2	4,2	0,7761	0,3114
Starting speed - 5m	s	1,14	1,16	1,12	0,5310	0,2307
Maximum speed - 10m	s	1,47	1,49	1,45	0,2880	0,0538
Explosive strength of lower limbs	cm	187,1	176,3	190,2	0,0273	0,0749
Explosive strength of upper limbs	cm	9,2	10,1	9,8	0,1573	0,1235
Agility	s	11,2	11,5	11,0	0,0368	0,0749
Endurance - shuttle run	stages	8,5	7,4	8,8	0,0022	0,0848
Suppleness	cm	20,4	20,5	22,5	0,6573	0,1108
PWC_{170}	W/kg	1,9	1,7	2,1	0,0104	0,0889
Basic technique - $W_{TP}1$	pts	27,8	30,2	32,1	0,0034	0,0315
Special motor skills - $W_{SUM}1$	s	64,3	66,4	62,4	0,0841	0,0885
Special aptitude - $W_{SS}1$	pts	43,4	45,6	51,6	0,0025	0,0080
Basic technique - $W_{TP}2$	pts	27,4	31,4	32,3	0,0007	0,0007
Special motor skills - W_{SUM} 2	s	66,7	69,3	63,8	0,0378	0,0101
Special aptitude - W_{SS} 2	pts	0,41	0,46	0,51	0,0010	0,0007
Coordinative special aptitudes -W_{KZS}	s	2,4	2,9	1,4	0,0891	0,0369
Synthetic special aptitude - W_{SSS}	pts	0,42	0,45	0,51	0,0013	0,0094
Intelligence (spatial imagination)	pts	30,6	32,9	34,3	0,5599	0,3710
Play effectiveness	%	42,8	40,6	50,9	0,0077	0,0885

When analyzing the mean values we note that the best characteristics show the athletes from Group 3. They include the shortest girls. They are characterized with higher maximum power output and total work, lower fatigue index, high PCW_{170}. They have the best basic technique, motor skills, special aptitude, coordinative and synthetic special aptitudes and the best game effectiveness.

Group 1 consists of the tallest girls with the worst basic technique.

Group 2 is characterized with higher level of fat content. These athletes have low values of explosive force of lower limbs, worst agility, low endurance and the worst motor skills.

If we calculate the mean synthetic indexes (computed from all features) for all groups, the greatest value is obtained for Group 3 - 52 and the lowest for Group 2 - 42. Group 1 has the synthetic score slightly better than Group 2, namely 44. The differences between group 1 and 2 are both structural and expressed as the mean level.

3. Factor analysis

Making the analysis of the collected material, at first the factorial structure of the investigated features has been investigated in the whole set of 396 observations, and then the structure of the features in separate groups. Because of the fact that the starting statistical material contained three measurements each from five age groups, it has been decided that the size of these groups are not so great as to preclude connecting this material into one common set of data, the more so that these considerations have also illustrative character. Of course, in the data set we are not dealing with independence, because each object supplied three measurements. It seems however, that it does not greatly interfere with testing interrelation, regression and factorial analysis. In the analysis of dynamics the compound data require special treatment, which will be shown later.

Using the factorial analysis for the set of compound data, we obtain six factors. This number results from the criterion saying that the characteristic value corresponding to the common factor cannot be smaller than one (i.e. the variance "carried" by the factor is not smaller than the variance corresponding to a single variable). The factorial analysis was effected on the primary correlation matrix with ones on the main diagonal.

Table 6. The characteristics of the factors

Factor	Characteristic value	Percentage of the interpreted variable	Cumulated percentage of the interpreted variable
1	10,93	42,0	42,0
2	2,62	10,1	52,1
3	1,59	6,1	58,2
4	1,35	5,2	63,4
5	1,26	4,8	68,3
6	1,21	4,7	72,9

After effecting the *normalized Varimax* rotation, we obtain the following factorial loads. In the following table the loads of absolute value above 0.6 are marked with shading. The identification of these variables enabled to name the factors.

Table 7. The factor loads

Variable	**Speed/ Force**	**Endur-ance**	**Power output**	**Tech-nique**	**Morp-hology**	**Fitness**
Body mass	0,057	-0,318	0,377	0,098	0,660	-0,134
Body height	0,069	-0,007	0,061	0,117	0,720	0,116
Fat content	-0,273	-0,560	0,210	0,054	0,505	-0,154
Maximum power	0,581	0,179	0,531	0,235	0,061	-0,052
Total work	0,552	0,132	0,584	0,290	0,132	0,027
Fatigue index	0,108	0,031	0,798	0,181	0,078	-0,332
Time to obtain maximum power	-0,117	0,018	-0,760	-0,293	-0,031	-0,343
Duration of maximum power	0,042	-0,019	-0,029	0,060	0,018	0,959
Starting speed - 5m	-0,770	0,027	0,006	-0,160	0,091	0,008
Maximum speed - 10m	-0,761	-0,028	-0,108	-0,204	0,134	-0,051
Explosive strength of lower limbs	0,769	0,128	0,353	0,251	0,117	0,084
Explosive strength of upper limbs	0,495	-0,055	0,380	0,178	0,420	0,093
Agility	-0,784	-0,233	-0,138	-0,186	-0,080	-0,077
Endurance - shuttle run	0,069	0,809	0,082	0,183	-0,083	0,032
Suppleness	0,067	-0,118	0,097	0,030	-0,650	0,026
PWC_{170}	0,079	0,756	0,053	0,013	0,112	-0,094
Basic technique - $W_{TP}1$	0,349	0,032	0,173	0,834	0,054	0,060
Special motor skills - $W_{SUM}1$	-0,601	-0,093	-0,134	-0,557	-0,314	-0,029
Special aptitude - $W_{SS}1$	0,477	0,055	0,176	0,809	0,149	0,038
Basic technique - $W_{TP}2$	0,361	0,016	0,188	0,827	0,056	0,006
Special motor skills - W_{SUM} 2	-0,661	-0,057	-0,092	-0,615	-0,238	-0,003
Special aptitude - W_{SS} 2	0,512	0,030	0,165	0,807	0,122	-0,012
Coordinative special aptitudes -	-0,578	0,042	0,034	-0,545	0,014	0,057
Synthetic special aptitude - W_{SSS}	0,503	0,042	0,172	0,820	0,136	0,012
Intelligence (spatial imagination)	-0,123	0,115	0,162	0,488	-0,073	0,054
Play effectiveness	0,621	0,084	0,002	0,398	-0,132	-0,100

The results of the factor analysis may be among others used for choosing the synthetic variables utilized for the construction of the synthetic index. If we choose from each factor one feature, having the greatest module of load (is most representative for the relevant factor), we will obtain the following list of features:

1. Agility
2. Endurance - shuttle run
3. Fatigue index
4. Index of basic technique in version I $W_{TP}I$
5. Body height
6. Duration of maximum power output.

The value of the synthetic index, constructed with the use of the above six features, was determined for the set of compound data. Fig. 3 shows the relation between the full index an the index using six features. In this figure, the continuous line plots the regression function and the dashed line marks the diagonal corresponding to the equal values of the indexes. It may seen that the correlation is strong and the index computed with the use of six features has values slightly smaller than the index computed with 26 features. This is not the result of smaller number of features, but their configuration.

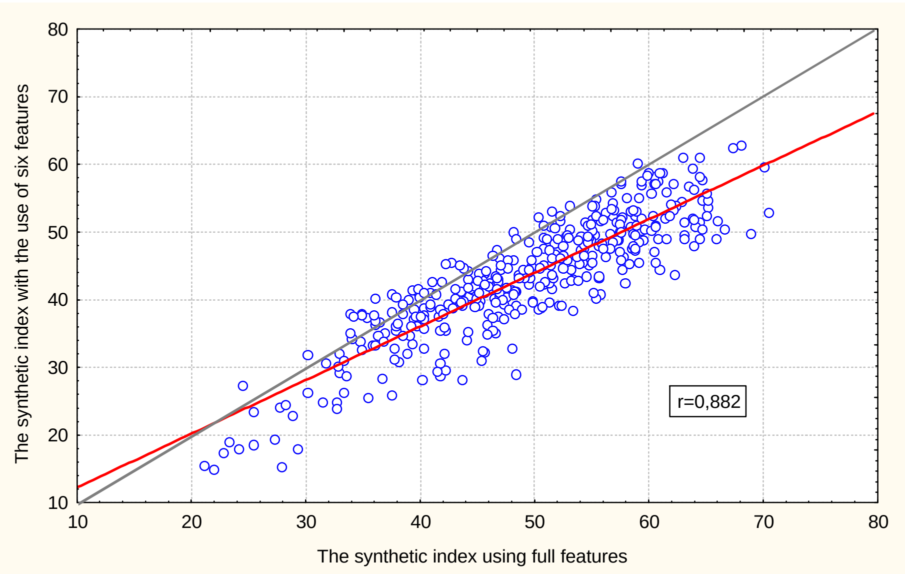

Fig. 3. The correlation of synthetic indexes

4. Discriminative analysis

The aim of the discriminative analysis is to find the computational principle enabling ordering object to classes of known number. For the correct use of this method in practice we need three data sets: teaching set, test set and investigated set. On the basis of the teaching set we verify the significance of proposed diagnostic variables and estimate the so-called classifying functions. On the objects of the teaching set we must of course know, to which subsets (classes) they belong. Such ordering is also known for the objects of the test set. Using this set, we check the effectiveness of the correct ordering with the use of the classifying functions, obtained from the teaching set. Finally, the investigated set is a set of elements, the ordering of which we do not know and want to classify them.

In our illustrative example we will demonstrate an attempt of using the discriminative analysis for the identification of the athletes with potential for improving their position in the training group. As a teaching set the athletes born in 1981 are used. On the basis of the ranks shown in Table 9, for each athlete an index of improvement of her relative position in the group was determined. calculated from the following formula:

$$I_i = (n+1-r_{Ci})*100/(n+1-r_{Ai}),$$

where:

n is the number of athletes in the group (it is also the greatest rank),

i is ordering number of the athlete,

r_{Ai} - the rank in the first measurement,

r_{Ci} - the rank in the third measurement

If the value of an index is greater than 100, this means that the athlete has improved her position in the group, and the value (I_i – 100) indicates the percentage of the improvement. The index smaller than 100 indicates that given athlete has decreased her position in the group.

In the estimation of the classifying functions the principle of stepwise elimination of the diagnostic variables is used. In the way similar to the descending stepwise regression, from the set of the diagnostic variables one by one variable of the highest p value is eliminated, until all p are not greater than the assumed level α.

Table 8 lists variables, corresponding p values and coefficients of the discriminative functions.

Table 8. The coefficients of the discriminative functions

Variable	*p* value	Coefficients of the discriminative function for the "no improvement" group	Coefficients of the discriminative function for the "improvement" group
Absolute term of the classifying function	-	-9560,61	-9448,69
Fat tissue	0,0062	-2,92	-1,79
Fatigue index	0,0032	-11,78	-12,70
Duration of maximum power output	0,0027	20,16	18,77
Explosive strength of lower limbs	0,0060	-1,10	-0,94
Suppleness	0,0410	7,19	6,97
Index of the basic technique in version I - $W_{TP}1$	0,0752	-649,15	-644,48
Index of special motor skills – $W_{SUM}1$	0,0581	296,00	293,76

Of course, for the estimation of the parameters of the discriminative functions, the values of the features from the first measurement were used. For this reason, the discriminative analysis may be used for predicting the relative development in sports training. At the same time, as can be seen, the features were identified, which are most significant for the development in the defined period of sports ontogenesis of teenage girl handball players.

For the classification of the athlete we have to compute for her the value of both discriminative functions. The greater result indicates to which group given person should be ordered. When the teaching set was classified, it was found out that 11 of 14 (i.e. 78.6%) of the athletes, who have improved their position, were correctly identified. In the group of athletes without improvement this percentage was decidedly higher. Of 24 girls, 22 (91.7%)

were correctly identified. This result is not surprising, as we are classifying a set which was used for the estimation of the classifying functions. Proper evaluation of the predictive qualities of the method may be obtained with classification of the test set. Among our data there is no set fully uniform with the teaching set. However, a set of measurements of girls born in 1982 was used here. It should be remembered that these athletes have different stage of development and training than girls born in 1981 and this may cause additional errors. After determination of the values of the classifying functions for the data of girls born in 1982, effecting the classification and comparison with actual changes of position, it was found that again the predictive value is significantly higher for the athletes, for whom no fast progress is expected, improving their relative position. Of 23 athletes who did not improve their position, the correct identification was made for 20 (87.0%). Of 10 athletes who did improve their position in the group, just a half were correctly identified. Once again, we stress the fact that the test group has different level of competitive development compared with the teaching group and the relations forming the base of the classifying functions could have changed.

Conclusions

As we have shown in the introduction, the phenomena connected with sports, especially with sports training, cannot be explained with single variables and randomly chosen tool of statistical analysis. The reduction approach, often based on the canons of J.S. Mill (Fisher 1971) is a thing of the past. The order of today is the multivariable and multidimensional analysis (Sokolowski 1998). Multidimensional character of scientific analyses could be interpreted in many ways. In this paper it is understood in terms of variable spaces and analytical tools.

Returning to the presented research problem and at the same time answering the first research question, it can be said that the W_i index, introduced by A. Sokolowski (1998), enables linear sorting of the investigated objects, and at the same time the evaluation of the changes on the scale of effects of the sports development of separate athletes and whole group. It has special usefulness in the longitudinal of sports training, because it demonstrates the individual differences in the flow of this process.

Investigating the population of the athletes we wanted to find their differences and similarities from the point of view of such property, as the differentiation caused by different reaction to the exercise loading. The application of the cluster analysis enabled the identification of the subgroups of the investigated girl handball players. The athletes have significantly differed in the scope of structural characteristics and special aptitude. The cluster analysis, based on variance analysis, has shown great usefulness in the typology of objects and analyzed features.

Interesting information was obtained with the factorial analysis. The obtained results justify the assumption that the analyzed features are complex grouping of aptitudes and confirm the conclusion of other workers (Raczek et al. 1998), indicating common conditioning of the motor effects, evaluating the level of anaerobic total work, maximum power output, explosive force of the legs, explosive force of the arms, start speed, maximum speed and agility.

The results of this work have shown that discriminative analysis may be used for predicting the relative progress in sports teaching. With the use of it the features were identified, which are most significant for the development in the specific period of sports ontogenesis of girl handball players.

In conclusion it may be said that the discussed tools of the statistical analysis are complementary but not substitutive.

REFERENCES

Campbell D. T. 1963. Experimental and Quasi-experimental Designs for Research. Rand MacNally

Dziembala L. 1975. The Principles of Statistics (In Polish). Akademia Ekonomiczna, Katowice.

Fisher R.A. 1971. The Design of Experiments. Hafner Press.

Frankfort-Nachmias C., Nachmias D. 1996. Research Methods in the Social Sciences. Martin’s College Publishing Group.

Grabinski T. 1992. The methods of taxonometry (In Polish). AE, Kraków.

Hellwig Z. 1968. The application of the taxonomical method to the typological division of the countries based on their level of development and resources and structure of qualified staff (In Polish, English summary). Przeglad Statystyczny, nr 4.

Hellwig Z. 1969. The problem of the optimal choice of predicants (In Polish). Przeglad Statystyczny, nr 3.

Raczek J., Mynarski W., Ljach W. 1998. Theoretical and empirical principles of developing and diagnosing of the motor aptitudes (In Polish, English summary). AWF, Katowice.

Rygula I. 2001. The tools of the system analysis of the sports training. AWF Katowice.

Rygula I., Jarzabek R. 2000. The diagnostic value of analytical tools in women's team handball. Biology of Sport.

Sokolowski A. 1998. The methods and measures of the comparison of the level and dynamics of the development of the regions. AE, Kraków.