World Academy of Science, Engineering and Technology
International Journal of Sport and Health Sciences
Vol:5, No:5, 2011

Multivariate Statistical Analysis of Decathlon Performance Results in Olympic Athletes (1988-2008)

Jaebum Park, and Vladimir M. Zatsiorsky

***Abstract*—**The performance results of the athletes competed in the 1988-2008 Olympic Games were analyzed (n = 166). The data were obtained from the IAAF official protocols. In the principal component analysis, the first three principal components explained 70% of the total variance. In the 1st principal component (with 43.1% of total variance explained) the largest factor loadings were for 100m (0.89), 400m (0.81), 110m hurdle run (0.76), and long jump (–0.72). This factor can be interpreted as the 'sprinting performance'. The loadings on the 2nd factor (15.3% of the total variance) presented a counter-intuitive throwing-jumping combination: the highest loadings were for throwing events (javelin throwing 0.76; shot put 0.74; and discus throwing 0.73) and also for jumping events (high jump 0.62; pole vaulting 0.58). On the 3rd factor (11.6% of total variance), the largest loading was for 1500 m running (0.88); all other loadings were below 0.4.

***Keywords*—**Decathlon, principal component analysis, Olympic Games, multivariate statistical analysis.

I. Introduction

In 1962, V. Zatsiorsky and M. Godik performed factor analysis (centroid method) of the decathlon performance in athletes participated in 1960 Olympic Games [6]. The study was motivated by the desire to determine a limited number of latent factors ('motor abilities') that define success in decathlon and, as a consequence, to help coaches and athletes in designing optimal training programs that take into consideration the inter-event similarity and possible transfer of training results. The authors analyzed the individual events as well as the overall performance (in awarded point scores). The factor loadings on the first factor were the largest for the total decathlon performance. For the individual events the loadings were almost identical to the magnitudes of their coefficients of correlation with the overall performance in decathlon. The factor was identified as the 'general level of athletic mastership' and hence the intended purpose of the research was not fully achieved. Including the overall performance in the factor analysis most probably masked the existing 'factor structure' of the decathlon events.

Van Damme *et al.* [5] approached the problem of interdependency between the decathlon events from a more general, biological, perspective. Their interest was in testing the so-called 'principle of allocation' which states that the physical performance by vertebrates is strongly influenced by tradeoffs between the pairs of ecologically important features and between conflicting specialist and generalist phenotypes [4]. The authors constructed a data set of the performance of 600 world-class decathletes. Similar to the previous findings [6] individual performance for any pairs of disciplines was positively correlated. The authors considered this fact 'unexpected'. According to their opinion, the above mentioned 'principle of allocation' should predict that high performance in one group of events is detrimental to the good performance in the disciplines with different requirements. For instance, speed events, such as 100 m dash, favor athletes with a large proportion of fast, fatigue sensitive muscle fibers while performance in endurance events, such as 1500 m running, relies on a larger proportion of slow, fatigue resistant fibers. When the authors restricted the sample to the top-notch athletes who scored more than 8000 points (n =133), they found negative correlation between the performances in the 100 m and 1500 m running (r = -0.021, p =0.016), as they expected. In the opinion of the authors, this fact confirms the 'principle of allocation'

Responding to the research of Van Damme *et al.* [5], Kenny *et al.* [3] analyzed the data of 92 decathletes participated in Olympic Games and found opposite results to that of Van Damme *et al.* They concluded that to compete successfully at the high level, a uniform high performance in all individual disciplines are required.

Neither Van Damme *et al.* [5] nor Kenny *et al.* [3] used multivariate statistical methods in their analyses. These methods were employed by Cox and Dunn [1] who used the cluster analysis and classical scaling to analyze the decathlon data collected at the meetings of the Amateur Athletic Federation from 1991 to 1999. The missing scores were replaced by their estimates obtained from linear regressions. The main conclusion was that the contemporary point awarding system favors those athletes who do well in field events.

The results obtained so far suggest that the outcome of the statistical analysis of the decathlon performance greatly depends on (a) the method used (for instance, using the centroid method of factor analysis, as it was done in [6] , is in our opinion suboptimal) and (b) the selected group of

Jaebum Park is with The Pennsylvania State University, University Park, PA 16803 USA (phone: 814-863-0354; fax: 814-863-4735; e-mail: jzp12@ psu.edu).

Vladimir M. Zatsiorsky is with The Pennsylvania State University, University Park, PA 16803 USA (phone: 814-865-3446; fax: 814-863-4735; e-mail: vxz1@ psu.edu).

athletes.

The goal of the present study is to repeat the study done by one of the authors 50 years ago [6] using the more powerful methods of analysis and the recent performance results.

II. Methods

The performance results of the decathlon athletes competed in the 1988-2008 Olympic Games were analyzed. The data were obtained from the IAAF official protocols [2]. The total number of athletes was 166 (1988: 34 athletes; 1992: 28 athletes; 1996: 28 athletes; 2000: 24 athletes; 2004: 28 athletes ; 2008: 24 athletes). Three athletes for whom some point scores were zero, evidently due to the disqualification, were excluded from the analysis. Table I summarizes some statistical data about the analyzed group of athletes.

TABLE I
THE DECATHLON PERFORMANCE RESULTS
(1988-2008 OLYMPIC GAMES, N = 166)

	Best Performance	Mean	SD	Unit
Points	8893.00	7978.36	505.09	
100m	10.44	11.04	0.28	sec
Long jump	8.07	7.23	0.35	meter
Shot put	16.97	14.32	1.29	meter
High jump	2.27	1.99	0.09	meter
400 m	46.41	49.33	1.31	sec
110 m hurdle	13.47	14,72	0.56	sec
Discus throwing	53.79	43.46	4.33	meter
Pole vaulting	5.70	4.74	0.35	meter
Javelin throwing	73.98	59.02	6.34	meter
1500 m	252.07	278.25	12.85	sec

The statistical analysis was performed using the SPSS software package (IBM Corporation, Somers, NY 10589). The principal component analysis (PCA) was used with the Kaiser Criterion (i.e., extract PCs with $\lambda > 1$) to extract the significant principal components (PCs). The rotation method was the Varimax with the Kaiser criterion. The rotations converged in 5 iterations.

III. Results

The obtained correlation matrix is presented in Table II. We did not test whether the estimates of the coefficients of correlation differ from zero because: (a) the data distribution is evidently not normal, (b) the studied group barely can be considered a 'random sample from a general population' as it is required by classical statistics for testing statistical hypotheses. In our opinion it is not a sample at all, there is no other Olympic athletes in the world except the studied ones. As a point of reference, we mention that for n=166 the estimates of r would be considered statistically significant at $p \leq 0.05$ if their values exceed 0.13.

With the exception of 1500 m running the magnitudes of all coefficients of correlation with the decathlon performance were not smaller than 0.6. The highest magnitudes of the coefficients of correlation were obtained for the long jump (r = -0.77) and the 110 m hurdle running (r = -0.75), the lowest for 1500 m running (r = -0.29). The lowest inter-discipline coefficients were obtained for 1500 m running. The only coefficient exceeding 0.35 was obtained with 400 m running. All other coefficients of correlation were below 0.3. It seems that participation in 1500 m running require from the athletes the traits that are not manifested in other disciplines. The highest inter-disciplinary coefficient of correlation was between the performance results in shut putting and discuss throwing (r = 0.78).

The PCA yielded three components with the significant contribution to the total variance (Table III). The first three principal components explained 70% of the total variance. Hence, a substantial proportion of the observed variance may be attributed to only three traits (latent factors).

TABLE II
CORRELATION COEFFICIENTS

	Points	100 m	Long jump	Shot put	High jump	400 m	110 m	Discus	Pole	Javelin	1500 m
Points	1.00	-.69	-.77	.72	.61	-.60	-.75	.65	.70	.60	-.29
100 m		1.00	-.65	-.50	-.26	.61	.70	-.38	-.37	-.19	-.02
Long jump			1.00	.40	.47	-.53	-.59	.35	.45	.32	-.20
Shot put				1.00	.39	.-20	-.49	.78	.41	.45	.11
High jump					1.00	-.21	-.35	.36	.34	.28	-.19
400 m						1.00	.52	-.16	-.35	-.13	.38
110 m hurdle,							1.00	-.41	-.46	-.34	.08
Discus throwing								1.00	.33	.38	.10
Pole vaulting									1.00	.40	-.16
Javelin throwing										1.00	-.11
1500 m											1.00

TABLE III
TOTAL VARIANCE EXPLAINED IN THE PCA: SUM OF SQUARED LOADINGS

Component	Total	% of variance	Cumulative %
1	8893.00	7978.36	505.09
2	10.44	11.04	0.28
3	8.07	7.23	0.35

The factor loadings on the extracted PCs are presented in Table IV. The data allow for a meaningful interpretation of the PCA results. In the 1st principal component (with 43.1% of total variance explained) the largest factor loadings were for 100 m (0.892), 400 m (0.814), 110 m hurdle run (0.764), and long jump (-0.718). This factor can be interpreted as the 'sprinting performance'. The loadings on the 2nd factor (15.3% of the total variance) presented a counter-intuitive throwing-jumping combination: the highest loadings were for throwing events (javelin throwing 0.758; shot put 0.743; and discus throwing 0.728) and also for jumping events (high jump 0.623; pole vaulting 0.567). On the 3rd factor (11.6% of total variance), the largest loading was for 1500 m running (0.881); all other loadings were below 0.4.

TABLE IV
FACTOR LOADINGS

Event	Components		
	1	2	3
100m	**0.892**	-0.178	-0.194
Long jump	**-0.718**	0.371	-0.167
Shot put	-0.338	**0.743**	0.379
High jump	-0.204	**0.623**	-0.243
400 m	**0.814**	-0.014	0.334
110 m hurdle	**0.764**	-0.360	-0.055
Discus throwing	-0.243	**0.728**	0.382
Pole vaulting	-0.373	**0.567**	-0.199
Javelin throwing	-0.032	**0.758**	-0.128
1500 m	0.113	-0.092	**0.881**

The largest loadings are boldfaced.

IV. DISCUSSION

As performance results in all individual events (with the exception of 1500 m running) have reasonably large correlations with the total number of awarded points we conclude that our data agree with the opinion of Kenny *et al.* [3] that to compete successfully at the high level uniform high performance in all individual disciplines (with the exception of maybe 1500 m run) is required, or at least highly desired. The correlation between the overall decathlon performance in all 10 events and performance in individual events is due to two sources: firstly, this is essentially a correlation between the 'whole' (total number of points) and the 'part' (performance in individual events); secondly, this correlation is affected by the inter-discipline correlations seen in Table I.

Among the three discovered factors, two —the 1st and the 3rd—allow for straightforward interpretation. They can be interpreted as the 'sprinting abilities' and 'endurance, respectively. The independence of these factors (Table 4) and a close to zero correlation (r = -0.02) between the achievements in 100 m and 1500 m running events suggest that the results in 1500 m run do not depend on the sprinting abilities of the athletes and, hence, 'maximal running speed' and 'endurance' (as it is manifested in 1500 m performance) are independent motor abilities. In this regard, our data confirm the concept advocated by Van Damme *et al.* [5] and the 'principle of allocation' [4] in general.

We tried to address the endurance problem in more detail. With this aim, we performed the regression analysis between the 400 m performance and the 100 m performance and determined the regression residuals. Our assumption was that these residuals represent the athlete's endurance in 400 m running. The logic was as follows: if two athletes A and B have the same performance results in 100 m dash, for instance 11.0 s, and different achievements in 400 m run, for instance, 48.0 s and 52.0 s, respectively, the different performance in 400 m in these athletes is due to their different endurance that can be parameterized by the regression residuals in the 400m *vs.* 100 m relation. Then we tested whether the 400 m endurance values (i.e., the above mentioned regression residuals) correlate with the 1500 m time. The answer was negative; the correlation was close to zero. Hence, if both 1500 m running time and the 400 m- vs.-100 m residuals are measures of the endurance they evidently represent different kinds of endurance, which are specific for 400 m and 1500 m running events. We also computed the regression residuals from the 1500/100 m and 1500m/400 m pairs and correlated all of them. The answer was also negative; the correlation coefficients were close to zero.

The 2nd factor (see Table IV) represented a counterintuitive grouping of the events: throwing and jumping effect together. The highest loadings were for the throwing events (javelin throwing 0.758; shot put 0.743; and discus throwing 0.728); the loadings for the jumping events were only slightly smaller (high jump 0.623; pole vaulting 0.567). It is a common experience, that throwers have usually large body weight while the jumpers (especially high jumpers) are tall but skinny. These traits are evidently advantageous for the performance in the above events. The best decathlon athletes manage to combine these controversial requirements and to achieve high performance results both in jumping and throwing events. For instance, the best performance results of the current World record holder Roman Šebrle (Czech Republic, 9,026 points, height 186 cm, weight 88 kg) are: in high jump 2.15 m, pole vaulting 5.20 m and in javelin throwing 71.18 m. All the results are quite impressive.

REFERENCES

[1] T.F. Cox TF, and R.T. Dunn, "An analysis of decathlon data," *The Statistician*, vol. 51, no. 2, pp. 179-187, 2002.
[2] International Amateur Athletic Federation official website. Available at www.iaaf.org.
[3] I.A. Kenny, D. Sprevak, C. Sharp, and C. Boreham C, "Determinants of success in the Olympic decathlon: Some statistical evidence", *Journal of Quantitative Analysis in Sports*, vol. 1, no. 1, Article 5, 2005.

[4] R. Levins (1968) *Evolution in Changing Environments*. Princeton University Press. New Jersey
[5] R. Van Damme, R.S. Wilson RS, B. Vanhoydonck, and P. Aerts, "Performance constraints in decathletes", *Nature*, vol. 415, pp. 755-756, 2002.
[6] V.M. Zatsiorsky, and M.A. Godik, "Mathematics and decathlon", *Track and Field*, vol. 10, pp. 28-30, 1962 (In Russian).