

Application and development of statistical methods and statistical-mathematical methods and models in the field of physical education in Serbia

Veroljub Stanković.

Faculty of Sports and Physical Education, the University of Priština
temporarily based in Leposavić, Serbia

e-mail: veroljub.stankovic@pr.ac.rs

1. Introduction

The beginnings of organized work effort and the formation of statistical service in our country date from the end of the 19th century (1862), and were primarily used for obtaining a census of the population and natural and economic resources. The work of these services led to the need for new research in all the fields of social life and work, and thus in the field of physical education. All types of knowledge, including the knowledge obtained in the field of physical education, are obtained by means of a scientific research method. Certain methods are used in this type of work, and researchers resort to certain courses. The methodology of scientific research work in physical education studies and systematizes these courses, methods, and research procedures, so as to be able to answer the important questions the researcher has to answer regarding his scientific-research work. Methodology in physical education can be defined as a part of physical education which deals with the evaluation of the means by which scientific knowl-

edge is obtained in this field. Statistics is a part of the internal structure of the methodology which evaluates the models, methods and algorithms for evaluation of data processing and hypothesis testing. The statistical methods which are used in physical education research can be found in the group of scientific disciplines which are directly or indirectly used in various forms of physical education, such as physical education classes, sport and recreational activities, or in other anthropological sciences. The very application and development of statistical methods in physical education are also closely related to the development of kinesmetrics, computer science and cybernetics.

2. Application and development of statistical methods and models in physical education

2. 1. Phase I

Statistical research during the 1950s and 60s in Serbia primarily took place in the Institute of Physical Education and Sports Medicine and also at various faculties of physical education, either as an independent scientific discipline or as part of other scientific disciplines. At the very beginning of the use of statistical methods and procedures it was primarily the features of a mass of participants which was being studied and whose units were classified, grouped into sequences with absolute or relative expressions of size and were shown in the form of tables and graphs. In the following phase we perform statistical analysis in which numerical data are obtained regarding the features of the analyzed occurrence and interpretation which leads to the possibility of drawing a conclusion. Generally speaking, during the first 20 years or so, the application of statistics in physical education in Serbia mainly began and ended with the following statistical procedures: taking a certain statistical group as a characteristic value marked as the parameter, its measure of central tendency or means were calculated, as well as its deviations from these means. The means are of different types and are selected according to group type: as the modus (M_o), the value which occurs most often, the median (M_d) which is in the middle of the distribution and the calculated means which include the arithmetic means (M), geometric means (G) and harmonic means (H). These means only partially characterize the group, since the significance of the greater or smaller dispersion is greater than the means. This dispersion indicates the homogeneity of the group and the extent of the confidence interval of the means. The following step includes determination of the dispersion measures such as the range of results, the variance, standard deviation, and the coefficient

of variance. After that we take into consideration the chronological sequence as the extent of the growth or the decline of a certain occurrence of the phenomenon over a certain period of time. The most frequent occurrences are the continuous developmental tendencies which are marked as a trend or as a line of movement during usually a longer period of time. It can either be linear increasing or decreasing, or curvilinear. The logical conclusion of this is the need to check all the individual influences in the movement of the occurrence and connection between certain variables in the movement of the occurrence. The extent of the connection is calculated through the correlation coefficient (r). If the degree of the connection between the occurrences is a complete correlation coefficient, it equals (1) and if the connection does not have a correlation coefficient, it equals (0). In physical education, we mostly come across incomplete (stochastic) connections, that is, the occurrence of a value of + or – for the correlation coefficient, or whether the increase in one variable follows the decrease in the other variable. Of the different correlation coefficients the ones most often used is Pearson's correlation coefficient (r_{xy}), rank correlation (ρ), the biserial coefficient (r_{bi}), and the contingency coefficient (C). After working on the previous problems of evaluation we move on to the comparison of the characteristics between the groups. On the comparison basis we determine whether any of the previously proposed hypotheses will be accepted or rejected, or in other words, we begin the hypothesis testing. This was carried out by means of the T-test, F-test and χ^2 -test. During the first phase of the development of statistical research in physical education, every procedure usually ended with the aforementioned statistical procedures.

2. 2. Phase II

We could say that the previous statistical procedures close off the first phase of the application of statistical methods in physical education in Serbia. Namely, research procedures in statistics were used for scientific purposes mostly regarding data which are quantified, that is, numerically expressed through measurement, while the measuring enables the use of statistical and thus mathematical methods. There is an opinion that the condition of a particular science can be evaluated by the extent to which its essence can be accounted for mathematically, that is, to which extent it could use math, even though mathematical methods are not the only ones which can be used in science. By observing the development of science we can see that the sciences whose subject matter is such that it facilitates the use of mathematical methods developed more quickly than those sciences in which it was difficult to use mathematics. The application of mathematics depends on the possibility of carrying out a measuring. Namely, where the dimensions of a particular occurrence can easily be abstracted from its whole and where they are

under relatively small influence of various factors, the measuring is easy, and vice versa. Towards the end of the 20th century, it was the use of statistical mathematical methods and models on the computer that opened an almost limitless field of possibility for deeper understanding of natural kinesiological processes and transformational processes which are used to form, shape, model, program, plan and follow them.

It is well known that the manifest anthropological space is irrelevant for the determination of the kinesiological laws for several reasons:

- Its dimensionality is virtually limitless,
- The manifested variables contain within themselves a specific variance and a variance of error,
- The interpretation of the manifested phenomena is either impossible or superficial and the stability of the results with the aim of their application during the prediction is significantly greater in the latent space than in the manifest space (Momirović, 1975).

The real kinesiological area is for these reasons only the space of latent kinesiological dimensions. Modern scientific methods, which include statistics as a method of scientific research, represent the scientific essence of the study of any phenomenon in physical education.

In physical education then, as well as now, scientific results have to enable the explication, prediction and transformation of the studied phenomena and processes in certain scientific disciplines which at the same time represent scientific laws as well. Numerous models exist for their operationalization, as well as methods, algorithms, and programs for the multivariate processing of data which are contained within various statistical packages, and are almost impossible to enumerate (Momirović et al. 1999). For the same reason their classification is limited to three classic scientific laws, or four different groups of methods which are most characteristic of kinesiological anthropological research (Malacko & Popović, 1997):

The methods for the determination of structure and classification:

- factor analysis,
- taxonomic analysis;

For determining influence and relations:

- regression analysis,
- canonical correlation analysis;

For determining differences and effects:

- multivariate analysis of variance and covariance,
- canonical discriminant analysis;

To determine changes and development:

- canonical analysis of change, and
- analysis of the development curve. (Popović, 1990, Malacko & Popović, 2001)

Even though the statistical-mathematical methods and models originate from the beginning and middle of the 20th century (the diagonal method of the factor extraction (Cholesky, 1915), the centroid method described by Thurstone (1935, 1947), the simple summation method described by Burt (1949), the canonical factor analysis by Lawley (1940), the varimax criterion suggested by Kaiser (1958), we did not need to wait long for their use in physical education (Stanković, 2000). To quote Štalec (2011), "real statistics only came into use from 1965 when professor Konstantin Momirović travelled from the Physical Education College in Zagreb to Ljubljana, and asked for a few correlational matrices to be calculated for what today is considered a small number of variables and participants (he remembers that there were 30 variables and 150 participants). Until then his associates had carried out analyses on hand-held calculators", data processing was partially done in Ljubljana under the guidance of professor Klas, and no computing center to be found in Yugoslavia in that day could (according to him) carry out those analyses on a computer. "Considering the fact that we successfully met these demands, the professor increased his appetite and he started designing more extensive studies and using various methods of multivariate data analysis". They programmed regression analyses, factor analyses, discriminant analyses and the multivariate multi-factor analysis of variance. This kind of prepared program was supposed to be fed into a computer, prepared in the form of a reading tape with the data on it and it was supposed to start the performance by means of a certain command on the computer stand. The program carried out command after command, no translation was necessary. Unfortunately not a single copy of the results can be found today.

In his overall lifework, dr Konstantin Momirović attempted to bring together Croatian and Serbian views and approaches as much as possible, acting as a fine connection between the two. The first paper written on physical education with the use of the statistical-mathematical method was written by dr Konstantin Momirović (1966) and it was entitled "Primjena latentnih antropometrijskih varijabli pri selekciji vrhunskih sportaša" (The use of latent anthropometric variables in the selection process of top athletes). Immediately afterwards he published his second article (1968) entitled "Faktorska struktura antropometrijskih varijabli" (The factor structure of anthropometric variables) published by the Kinesiology Institute of the Faculty of Physical Education in Zagreb, where he was working at the time.

In his PhD thesis entitled "Struktura morfoloških, motoričkih, kognitivnih i konativnih dimenzija u dizača tegova" (The structure of morphological, motor, cognitive and conative dimensions among weight lifters, 1976), which was defended at the Faculty of Physical Education in Belgrade, he also applied a multivariate statistical method including factor and regression analyses.

In 1979 dr Dragan Popović started working in Niš on the application and development of the DRSOFT program package for home and personal computers which work with CP/M, PS-DOS, MS-DOS and XENIX operative systems, and which deal with quantitative data processing. All the programs and sub-programs that the professor wrote in Fortran and Basic were tested over several dozen MSc theses and PhD theses and their descriptions were explicitly stated in the handbook "Programi i potprogrami za analizu kvantitativnih promena" (The programs and sub-programs for the analysis of quantitative changes) (1993) at the Center for Multidisciplinary research of the Faculty of Physical Education in Priština established in 1992, whose head professor Popović has been ever since it was established. Today we can say that it was a pioneer endeavor since, as far as the author was able to determine, at the time no one was writing statistical programs for these types of computers. Computing centers up till then were engaged in adapting program packages which were written for the IBM 1130 computers (the well-known SSP package) as well as for other computers which were installed throughout Yugoslavia at the time (PDP and the greatest computer installed up till then, the UNIVAC computer at the SRCE center in Zagreb. The first pioneer papers by professor Momirović and professor Popović were a modification of the algorithms and programs written in FORTRANU 4 of the well-known handbook by Cooley & Lohnes (1971) "Multivariate data analysis".

2. 3. Phase III

In a great number of research projects, the solution to the research issue in physical education could not be found in the closed algebraic form and because of that the solution had to be sought out in some numerical algorithms. As it is well known, up to now several hundred of these algorithms which were efficient enough to be retained in practice were proposed, and thus found their place in textbooks; several dozen of these algorithms were implemented in the most frequently used statistical program systems or packages such as SAS, GENSTAT, SPSS, Statistica, BMDP, SYSTAT and others, and were not necessarily worse program products. The use of these products in different sciences or fields derived from these sciences have shown that their effectiveness varies significantly not only from method to method, and from implementation to implementation, but also from problem to problem, defined by the configuration of the vectors of

the objects in the variable space. In such situations it is natural that the solution to the problem could be found in the space of heuristic methods and thus in the space of neural networks, for the simple reason that neural networks can be used to emulate any method of data analysis. These networks have been constructed in such an enormous number up to date, that it is virtually impossible to count them. Nevertheless, they are not all equally accessible, nor equally effective, and have not been checked in a rigorous enough fashion. With the arrival of professor Momirović to Belgrade (1991), Serbia became one of the world's centers for the development of neural networks, which in addition to the other social sciences found their use in kinesiology. Professor Momirović developed the following program packages: SS, GENSTAT and SPSS and wrote several hundred algorithms and programs for them. In collaboration, Momirović & Popović 2003, wrote one of the seminal works of our field which refers to neural networks "Konstrukcija i primena taksonomskih neuronskih mreža" (The construction and application of taxonomic neural networks). Professor Momirović and professor Popović developed networks which were implemented in the SPSS environment until 2003, and for a very simple reason, the fact that the SPSS environment is, on the one hand, easy to use and it requires no knowledge of programming and, on the other hand, is available to all the users of statistical packages. It is precisely for this reason that there are such great limitations and possibilities for the serious use of neural networks. That's why professor Popović decided on the implementation of the neural network in the SAS environment. There were two important reasons. The first reason is that the SAS environment has been designed in a much more serious form and of itself can offer the user far more, while the second reason is that it is the only environment which can fully satisfy all users and especially those who are knowledgeable regarding the production of algorithms and programs.

3. Conclusion

The beginnings of the organized work effort and the formation of a department for statistical analysis in Serbia date from the end of the 19th century (1862), which had the aim of completing a population census, as well as a census of natural and economic resources. At the very beginning when these statistical methods and procedures were being used for the first time, they were mostly descriptive statistics and tests statistics. After the second phase which is characterized by a long period of the use of statistical-mathematical methods for determining the structure, differences, the influence and relations within and between the analyzed anthropological spaces, we come to the phase of application of neural networks by means of which we can emulate any method of data analysis. These

networks are so numerous nowadays that it is practically impossible to count them all. Nevertheless, they are not all equally accessible, are not equally effective and they have not all been evaluated in a rigorous enough fashion. Further development of anthropological sciences (kinesiology, psychology, genetics, pedagogy, sociology, etc.) can necessarily be expected to be promoted if information technology improves, including the use of mathematic and cybernetic methods in the application and development of neural networks.

Lately there has been much abuse of the statistical packages and even terror in the so-called field of data processing. Having worked all these years with professor Popović, who is the only true follower of professor Momirović and his heir, first as an assistant and now as a professor myself, and having spent most of my time with him in the field of scientific research work, I have learned that it is impossible to accept the widespread belief that it is enough to come into possession of computer equipment, to obtain the cheapest statistical package and with a little good will carry out any serious scientific research. We would not be quoting and citing the names of certain authors who designed the name "data processors" if they did not in fact understand the basic logic of some of the multivariate analyses, or, for instance, hyper and hypo factorization in the factor analysis. What are we to do: either the modern young researchers of today will learn the basics of programming and the logic of multivariate models which have explicitly been covered to date in the numerous textbooks by many of our co-authors, so that they could use some of the more serious program packages, for instance, SAS and neural networks, or this hard work will all be in vain.

4. References

- [1.] Cooley, W. W. & Lohnes, P. R. (1971). *Multivariate data analysis*. New York, etc: Wiley.
- [2.] Kurelić, N., Momirović, K., Stojanović, M., Šturm, J., Radojević, Đ. & Viskiće-Štalec, N. (1975). *Struktura i razvoj morfoloških i motoričkih dimenzija omladine*. Beograd: Institut za naučna istraživanja Fakulteta fizičkog vaspitanja.
- [3.] Малацко, Ј. (1970). Валидација инструмената за испитивање психосоматског стања и степена тренираности рвача. Одобрен и реализован научни пројекат. Београд: Савез за физичку културу Југославије.
- [4.] Малацко, Ј. (1976). Структура морфолошких, моторичких, когнитивних и конативних димензија у дизача тегова.

- Необјављена докторска дисертација. Београд: Факултет за физичко васпитање.
- [5.] Malacko, J (1994). Антрополошка статистичка методологија. Нови Сад: СИА.
- [6.] Malacko, J. & Popović, D. (1997). Metodologija kineziološko antropoloških istraživanja. Приштина: Универзитет у Приштини, Факултет за физичку културу.
- [7.] Momirović, K. (1966). Primjena latentnih antropometrijskih varijabli pri selekciji vrhunskih sportaša. Zagreb: Institut za kineziologiju.
- [8.] Momirović, K. (1968). Faktorska struktura antropometrijskih varijabli. Zagreb: Institut za kineziologiju.
- [9.] Momirović, K. (1975). Metode za transformaciju i kondenzaciju kinezioloških informacija. Zagreb: Fakultet za fizičku kulturu.
- [10.] Momirović, K., Wolf, B. & Popović, D. (1999): Uvod u teoriju merenja 1, interne metrijske karakteristike kompozitnih mernih instrumenata. Приштина: Факултет за физичку културу.
- [11.] Momirović, K. & Popović, D. (2003). Konstrukcija i primena taksonomskih neuronskih mreža. Leposavić: Univerzitet u Приштини.
- [12.] Popović, D. (1990). Metodologija istraživanja u fizičkoj kulturi. Niš: Naučni podmladak.
- [13.] Popović, D. (1993). Programi i potprogrami za analizu kvantitativnih promena. Приштина: Факултет за физичку културу.
- [14.] Stanković, V. (2000). Testiranje i primena programa za utvrđivanje strukture i razlika u međusobno povezanim prostorima kod rukometaša u toku jednogodišnjeg trenažnog procesa. Neobjavljena doktorska disertacija. Приштина: Факултет за физичку културу.

Received on June 18th, 2012.
Approved on Sep. 19th, 2012.

Application and development of statistical methods and statistical-mathematical methods and models in the field of physical education in Serbia

Veroljub Stanković

*Faculty of Sports and Physical Education, the University of Priština
temporarily based in Leposavić, Serbia*

e-mail: veroljub.stankovic@pr.ac.rs

Summary

The beginnings of an organized work effort and the formation of a department for statistical analysis in our country date from the end of the 19th century (1862), and were primarily used for obtaining a census of the population and natural and economic resources. The work of these services led to the need for new research in all the fields of social life and work, and thus in the field of physical education. Modern scientific methods, such as statistics as a scientific method of research, represent the scientific basis of the study of any phenomenon in physical education. The statistical methods applied to physical education research can also be used in a group of related scientific disciplines which were used, directly or indirectly, in various forms of physical education in general, physical education in the curriculum, in sport and recreational activities. At the very beginning, the statistical method and procedures used were primarily those that evaluated the features of a studied group, whose units were classified and grouped into sequences with absolute or relative expressions of size, and which were presented in

the form of tables and graphs. After the second phase, characterized by extended periods of use of statistical methods to determine the structure, differences, influence and relations within and between the analyzed anthropological spaces, we come to the current phase characterized by the use of neural networks for the simple reason that neural networks can emulate any method of data analysis. Many of these networks have been constructed up to date, so many in fact that their number practically cannot be determined. Among them there is a definite, although not big, number of taxonomic neural networks. Nevertheless, they are not all equally available, not all equally effective, and have not all been evaluated in a rigorous manner.

Key words: statistical methods, statistical models, neural networks, historical development, Serbia

Correspondence:

Faculty of Sports and Physical Education, the University of Priština temporarily based in Leposavić, Serbia

Dositej Obradović st. nn

38218 Leposavić, Serbia

Tel.: ++ 381 28 84 701; ++ 381 64 8904600

E-mail: veroljub.stankovic@pr.ac.rs