SIMILARITY ESTIMATION OF TIME SERIES WITH SMALL DIMENSIONALITY

Antonova I.V., Chikina N.O., Soloshenko E.M.
National Technical University «Kharkiv Polytechnic Institute», Kharkiv
The State Establishment «Institute of Dermatology and Venereology of the National Academy of Medical Sciences of Ukraine», Kharkiv

When analysing time series describing the dynamics of the same statistical indicator characterising the state of the studying system, the task of assessing the similarity of these series is topical. The assessment is carried out in order to distinguish among them groups with «similar» indicator dynamics.

Most of the existing methods for estimating the similarity of time series are based on reducing their dimensionality, which is undesirable when analysing short time series. The actuality of this task is that statistics on many economic, social, demographic, environmental indicators are collected, as a rule, once a year over several years. In addition, access to this kind of information is very limited. This raises the question of how to present information about a time series of small dimensions so that it is a sufficiently accurate description of it.

One such method is the MPOX method. This method is based on the calculation of time series characteristics, which are divided into three groups, time-independent, describing dynamics and defining variability. For each group, a value reflecting the difference in time series characteristics from a particular group is calculated and then a linear combination of them is constructed.

In a joint study with the State Establishment «Institute of Dermatology and Venereology of the National Academy of Medical Sciences of Ukraine» (IDVNAMNU), the authors collected information on the incidence of allergic and various skin diseases among workers depending on their length of service in one of the chemical-pharmaceutical industries (Kharkiv).

To analyse this information, risk groups (GR) were formed as follows: 1) GR1 comprised workers who had a history of allergic reactions and had worked at the company; 2) GR2 included workers with various somatic diseases; 3) GR3 included workers with various skin diseases of non-allergic genesis; 4) a group of workers whose condition was defined as almost healthy (AH).

Since research was carried out on the development of allergies and various skin diseases in workers depending on their length of service, the information on the number of workers in the risk groups varied almost every year. Following a time-series similarity assessment algorithm, a number of patterns were identified in the dynamics of the number of workers in each risk group.

The biggest change was in GR1. The size of this group decreased over time, as the high risk of morbidity forced workers to change jobs. There was little or no change in the AH group after 12 years of service. The highest time-series similarity values were found in the GR2 and GR3 series.