UDC 004.04

APPLICATION OF SAS FOR ANALYSIS OF DATA OF MEDICAL INSTITUTIONS FROM DATA.EGOV.KZ PORTAL

Zh.R. Amanzholova[1], Z.A. Altayeva[2], A.N. Moldagulova[3] L.M. Alimzhanova[4] A.A. Kuatbayeva[5]
International IT University (IITU), Almaty, Kazakhstan

[1]https://orcid.org/0000-0002-6866-7667, amanzholova.97@mail.ru
[2]zarina_altayeva@mail.ru, [3]a.moldagulova@iitu.kz, [4]dimkim_01@mail.ru,
[5]ahamala2020@gmail.com

Abstract. The article discusses the statistical analysis of medical data on the example of infectious diseases using SAS. The analysis is based on a unified method for calculating indicators included in the State Program for the Development of Healthcare of the Republic of Kazakhstan "Densaulyk". The initial data for the research were taken from the open source data.egov.kz, which presents the latest information on infectious diseases in 3 regions and 1 city of the Republic of Kazakhstan. The methodology is based on the use of methods of descriptive statistics, graphical data analysis, building models on the SAS platform and can be used for a more in-depth analysis of infectious diseases in various population groups, as well as for their description in the cities of Kazakhstan. During the research, statistics were obtained using the SAS system for the level of infectious morbidity and the downloaded data was visualized in the form of various diagrams. The information obtained on the level and nature of infectious diseases can be used to assess trends in the health status of the population, plan various types of specialized medical care and treatment and preventive measures, and make rational use of the material and human resources of the healthcare system.

Keywords: data analysis, SAS platform, open source, infectious diseases, model.

Introduction

Health system management is not possible without a detailed statistical analysis of data from medical institutions. Currently, control over the activities of the system of a medical institution is largely based on statistical forms summarizing data on morbidity and mortality of the population, as well as characterizing the activities of medical institutions. Such an approach is generally based on the understanding of health as a state of absence of diseases, which runs counter to the modern definition of World Health Organization (WHO), according to which full physical, mental and social well-being is also necessary for full health [1]. And this entails the need to review the set of indicators necessary for the analysis of the epidemiological picture in the territory of administrative units of any level. Macrodata allows us to identify general trends, but they do not provide an opportunity to track the causes of the changes. Moreover, they do not reflect the relationships that exist between infectious diseases. All this determines the relevance of this research.

Since there is no access to more detailed statistical data, it was decided to analyze the data of patients from medical institutions of Akmola, Aktobe, Almaty regions and the city of Almaty from an open source data.egov.kz portal. Based on the analysis, it is possible to show the potential of microdata and their advantage over traditional macrostatistical forms. It is the characteristics of infectious morbidity and mortality at the municipal level that remain practically outside the interests of medical and demographic statistics, while at this level, the prevention of morbidity can become most successful, as it is as close as possible to individuals with their special problems [2].

The main goal of this research was to study the state of the source of open data from the data.egov.kz portal according to the data of medical institutions and the system of statistical registration of infectious morbidity and mortality at the level of medical institutions for its suitability for statistical and analytical studies. The novelty of the research lies in the development of a model for statistical analysis of data from medical institutions obtained from

an open source. The paper shows the methods of working with open data sources and a model for analyzing the epidemiological picture in the territories served by these medical institutions using the Statistical Analysis System (SAS) platform tools.

Prospects for the application of sas for analysis of data of medical institutions

The progress of modern medical science is based on the analysis of large volumes of fact evidence of clinical, laboratory and other research methods [3]. An objective assessment of these data is greatly facilitated by the methods of statistical analysis, which help to find out not only quantitative, but also qualitative relationships between the studied phenomena [4].

Recently, for the processing, analysis and presentation of data from medical institutions, many statistical packages have been used - one of which is the SAS system. The system consists of modules, each of which performs a certain range of tasks, and it also implements its own programming language. SAS offers analytical solutions that provide effective monitoring. SAS is used for data processing, as well as for educational purposes, to develop skills in students working with this program. Using SAS, it is possible to carry out statistical processing of data of various levels of complexity, in accordance with the tasks: dispersion analysis, multiple linear regression analysis, logistic regression, survival analysis, etc. The SAS system also has great capabilities for presenting data: creating graphs, charts, tables, both for publication and presentation purposes. Interaction with the program is possible both in console mode and through a graphical interface, which is a graphical shell for simplified input of SAS programming language commands [5].

The software market continues to grow. New software packages are developed on the basis of modern computer technologies, which are developing rapidly. The capabilities of the programs are expanding, allowing a person to use them not only as an improvised tool that facilitates work, but also as a full-fledged assistant capable of solving complex medical problems. This research shows the use of the SAS statistical package to analyze data from medical facilities.

Data processing and methods

Infectious morbidity is one of the main medical and statistical indicators of the health status of the population [6]. An analysis of this indicator over a number of years allows us to draw conclusions about the incidence and dynamics of the incidence, as well as the effectiveness of the complex of socio-hygienic and therapeutic measures aimed at reducing it.

The study was based on data on infectious morbidity and mortality in the Akmola, Aktobe, Almaty regions and Almaty city. The data describing the infectious morbidity in 2020 were selected as initial data. Cases of infectious diseases were recorded monthly. The diseases of the following diagnostic groups are presented: A00-B99 «Some infectious and parasitic diseases». When analyzing the infectious morbidity of the population, three main types of tasks were solved: comparing morbidity indicators in age groups, analyzing morbidity dynamics and analyzing morbidity.

In total, in 2020, 588 cases of infectious diseases were recorded in the Akmola region, in the Aktobe region - 1111, in the Almaty region - 1715, in Almaty - 334.

About 47.2% of patients are women and 52.8% of patients are men. The total number of appeals by territory is presented in Figure 1.

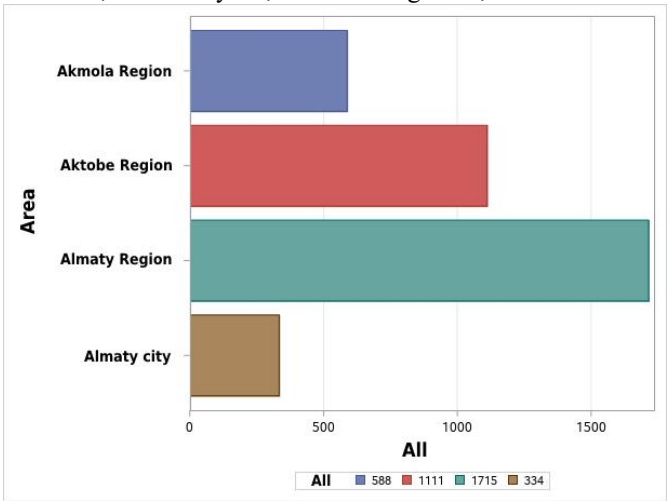

Figure 1 – Chart of the total number of appeals by territory

Results and analysis

The data of infectious diseases among the population in the Republic of Kazakhstan is analyzed and the growth dynamics of the disease by some infectious and parasitic diseases among the population is noted. In addition, indicators were described and analyzed that describe the incidence (in terms of population circulation) of the main age groups of the population (children, teenagers, adults) in the Republic of Kazakhstan with infectious diseases.

For comparison, the age groups in Figure 2 were taken: children (0-14 years old), teenagers (15-17 years old) and adults (18 years and older). The study noted an increase in the incidence rate among adults in early 2020. In children, the incidence rate is slightly lower than in adults. A significant decrease in the incidence rate among teenagers is also noted.

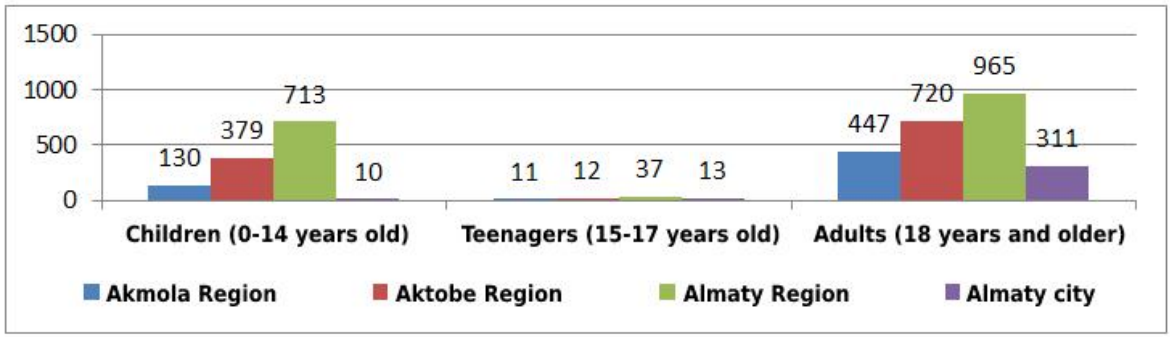

Figure 2 – Diagram of growth dynamics of infectious diseases of different age groups by region

During the processing of data on infectious diseases, the following parameters were taken to obtain statistics:

Distribution by regions;

The number of all patients with infectious diseases.

The diagram shows the statistics on infectious diseases of all patients in the Akmola, Aktobe and Almaty regions and in Almaty city (Figure 3).

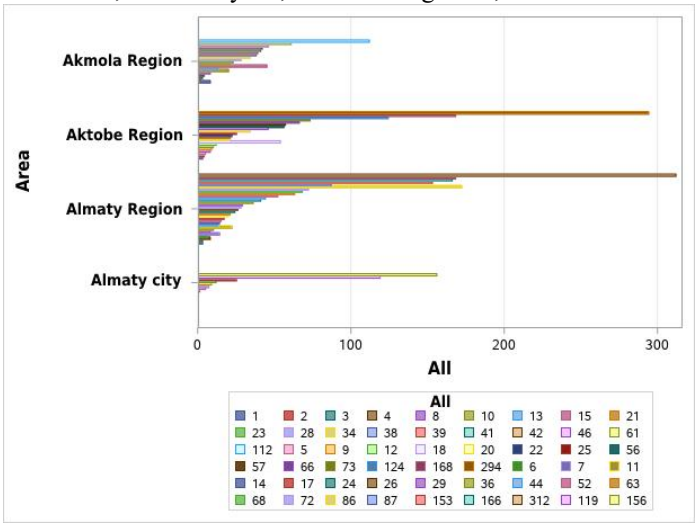

Figure 3 – Chart on infectious diseases of all patients by regions

The following statistics is given for the number of deaths from infectious diseases in the Akmola, Aktobe and Almaty regions and Almaty city (Figure 4).

Selected options:

Distribution by regions;

The number of deaths from infectious diseases.

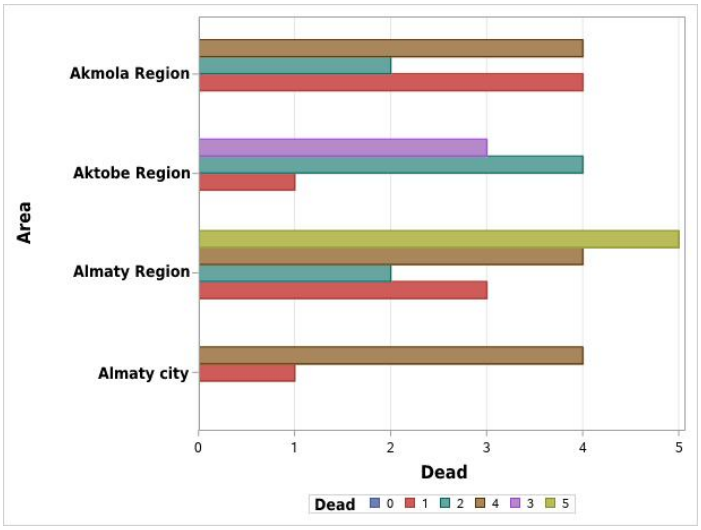

Figure 4 – Chart by the number of deaths from infectious diseases by regions

The following are statistics on the number of patients with certain infectious and parasitic diseases in the Akmola, Aktobe and Almaty regions and Almaty city (Figure 5).

Selected options:

Distribution by regions;

The number of patients with certain infectious and parasitic diseases.

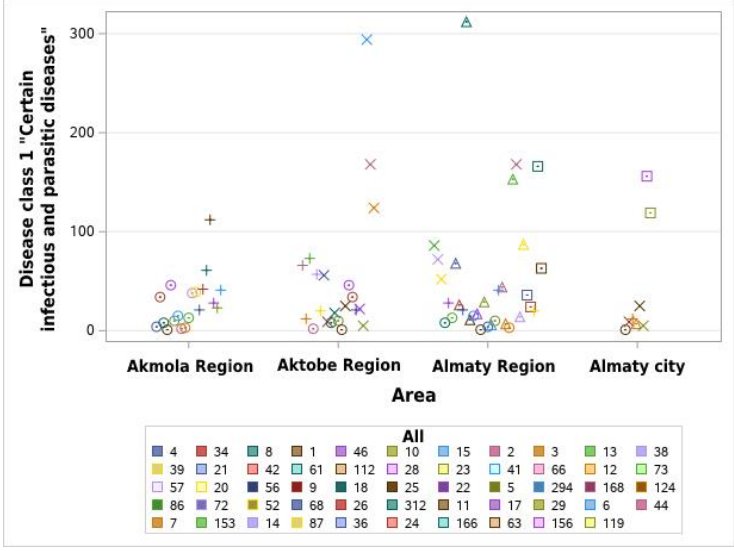

Figure 5 – Chart by the number of patients with some infectious and parasitic diseases by regions

Conclusion

In conclusion, it should be noted that a detailed analysis and processing of statistical information is an important component in the management of medical institutions. It is proposed to use statistical packages for processing, analysis and presentation of data from medical institutions - one of which is the SAS system. Using the SAS system, statistical processing of data of various complexity levels is carried out in accordance with the tasks: analysis of variance, logistic regression, survival analysis, etc.

The article uses the SAS system to analyze and process the data of medical institutions from the data.egov.kz portal on the level of infectious morbidity in the population of the Republic of Kazakhstan. Thus, in the course of the study, statistics was obtained using the SAS system for the level of infectious diseases for the beginning of 2020 and the downloaded data was visualized in the form of various diagrams.

As a result, statistics on the level of infectious diseases in the regions (Akmola, Aktobe, Almaty) and Almaty city were obtained:

For different age groups;

By the number of all patients;

By the number of dead;

By the number of patients with certain infectious and parasitic diseases.

The proposed methodology for studying the dynamics of infectious diseases can be applied for a more in-depth analysis of infectious diseases in various population groups, as well as for describing infectious diseases in other regions and cities of the Republic of Kazakhstan, presented in the open source data.egov.kz.

The information obtained on the level and nature of infectious diseases can be used to assess trends in the health status of the population, plan various types of specialized medical care and treatment and preventive measures, and make rational use of the material and human resources of the healthcare system.

References

[1] Ustav VOZ [Elektronnyy resurs]. Rezhim dostupa: https://www.who.int/ru/about/who-we-are/constitution (7 iyunya 2020)

[2] Jose Bien R. Hernandez, Peggy Y. Kim Epidemiology Morbidity And Mortality [Elektronnyy resurs]. Rezhim dostupa: https://www.ncbi.nlm.nih.gov/books/NBK547668/ (11 yanvarya 2021)

[3] National Academies of Sciences, Engineering, and Medicine. 2015. Improving diagnosis in health care. Washington, DC: The National Academies Press [Elektronnyy resurs]. Rezhim dostupa: https://www.nap.edu/read/21794/chapter/1#iii (11 yanvarya 2021)

[4] M.K. Iztleuov, G.G. Yerimbetova Ispol'zovaniye sistemy SAS dlya analiza rezul'tatov biomeditsinskikh issledovaniy. Batys K¸azak¸stan meditsina zhurnaly. 2012. 3(35). 12-19.

[5] Obzor statisticheskikh programm [Elektronnyy resurs]. Rezhim dostupa: http://www.sciencefiles.ru/section/46/ (7 iyunya 2020)

[6] Morbidity statistics in the EU Report on pilot studies // Publication Office of the European Union, Luxembourg 2014 [Elektronnyy resurs]. Rezhim dostupa: https://ec.europa.eu/eurostat/documents/3888793/5858521/KS-TC-14-003-EN-N.pdf/bd959e6e-10ed-4078-915e-308941c02811 (11 yanvarya 2021)