Talalaev K. O., Hoydyk V. S., Vastyanov R. S., Todorova A. V. HIV/AIDS statistical analysis and morbidity prediction among injection drug users in the Black Sea region of Ukraine. Journal of Education, Health and Sport. 2019;9(7):805-820. eISSN 2391-8306. DOI http://dx.doi.org/10.5281/zenodo.3571360
http://ojs.ukw.edu.pl/index.php/johs/article/view/7645

The journal has had 7 points in Ministry of Science and Higher Education parametric evaluation. Part B item 1223 (26/01/2017).
1223 Journal of Education, Health and Sport eISSN 2391-8306 7

The authors declare that there is no conflict of interests regarding the publication of this paper.
Received: 16.07.2019. Revised: 23.07.2019. Accepted: 30.07.2019.

UDC 616.98:578.828.6]-056.83-36.22/8(477.7)

HIV/AIDS STATISTICAL ANALYSIS AND MORBIDITY PREDICTION AMONG INJECTION DRUG USERS IN THE BLACK SEA REGION OF UKRAINE

K. O. Talalaev, V. S. Hoydyk, R. S. Vastyanov, A. V. Todorova

Odessa National Medical University, Odessa, Ukraine

Abstract

The development of methods for statistical analysis of large data volumes and availability of standard information processing packages to users shows signs of widespread their usage, including for medical forecasting. It is a scientifically sound prediction of possible future changes in public health in general or the health status of specific patients.

In order to make medical decisions and obtain information about public health in general or the health of individual patients the prediction of possible future changes is important. Different statistical approaches should be applied to validate the projected risks.

Comparatively two study groups tested for normality a number of criteria presented for analysis and concerning the incidence of human immunodeficiency virus/acquired immunodeficiency syndrome of residents of Odessa, Mykolaiv, Kherson regions and Ukraine as a whole, as well as injecting drug users, injecting drug users specified areas.

Using the modern methods of a statistical analysis the authors come to the idea that these absolute and relative data allow to make a complex scheme of HIV/AIDS prediction on the Black Sea region of Ukraine. Therefore, it was shown that Odessa and Kherson regions of the Black Sea region demonstrated a steady increase in the HIV incidence; the increase of the incidence makes around 10-20% annually. In the Mykolaiv region, on the contrary, there is an annual reduction in incidence, although this trend weakened in recent years. The nature of fluctuations in the incidence of HIV in the period from 2006 to 2018 differs in Odesa, Mykolaiv, Kherson regions and in Ukraine in general. The correlation between Odesa and Mykolaiv region – is negative, between Odessa and Kherson - almost zero between Ukraine and Odessa region – weakly positive

Key words: statistical forecast; mathematical based prediction; public health; HIV/AIDS; drug usage; Black Sea region of Ukraine.

Introduction. Statistical analysis of empirical data shows that the development of methods and tools for evidence-based medicine has promoted the computerization of the mass of biomedical research and the availability of standard data packages to users led to their wide application. In particular, probabilistic-statistical or predictive methods that were used to effectively correct and accurately describe the phenomena of massive industrial and financial sectors are now widely used in medical practice. In order to make medical decisions and obtain information about public health in general or the health of individual patients the prediction of possible future changes is important. Risk prediction models or the situation development as a whole can be created using different statistical approaches [1].

Sometimes it is often not taken into account that the complexity of solving data in medicine is not only the lack of reliable data on the distribution laws of statistical characteristics of random factors, but in the absence of a single universal algorithm formalization of such problems as a result of synergistic action of some conditional and unconditional factors. In modern conditions of pre-clinical diagnosis basic algorithms that significantly differ from current clinical approach should be relevant. This will apply the structural components of the first level of care or family doctors to target mass predictive surveys. Predictive (prognostic) diagnosis is considered to be the basis for targeted preventive measures and as a result of an individualized approach to the future of health and effective treatment, which is a more productive way of interaction between patient and doctor. [2].

The above factors and their interaction not only significantly slow down the process of formalization and finding adequate models of analysis of empirical data, but also have a significant and non-linear effect on the results of formation of control levers on the situation as a whole, i.e. appropriate management decisions in health care.

Some researchers, having processed a small amount of data and received some results of calculations forget that the inspiring "simplicity and obviousness" of programming possibilities for solving practical health is wrong and leads to the incorrectness of the estimates and judgments in terms of the theory of partial differential equations Zh.S. Hadamard derivatives [3].

Often this is due to the neglection of the fundamental need to verify the possibility of preconditions of probabilistic and statistical methods, especially when dealing with censored samples.

In this current case, having followed the basic tenets of mathematical statistics, a more detailed analysis of the data presented in the paper was conducted.

The aim of our study. For successful statistical analysis when comparing quantitatively two investigated groups and having the reasonable assumption about the homoscedasticity of available observations, to check the normality of a number of criteria that were submitted for analysis, the raw data regarding the incidence of human immunodeficiency virus (HIV)/acquired immunodeficiency syndrome (AIDS) in residents of Odessa, Mykolaiv, Kherson regions and Ukraine in general and injection drug users (IDU's) that live in these areas.

Results

1. HIV/AIDS distribution in Odessa region analysis

Figure 1 reflects a histogram distribution for the incidence of AIDS among IDUs, and Figure 2 – for the overall incidence rate in the Odessa region.

The null hypothesis test about the fact that the samples are subject to normal distribution, was in choosing the appropriate criteria. For the investigated parametric case, according to the amount of data the Shapiro-Wilk criterion was selected [4]. The reliability of the analysis in this case is confirmed by the high rate of Shapiro-Wilk criterion significance. It is also confirmed by the high value of $p>0.05$ [5], which allows for parametric methods of statistics.

When analyzing Fig. 1 it was determined that the resulting histogram distribution of IDU patients is quite compact and has a slight variance. The presence of two events clearly shows the composition of distribution laws, but it is not essential for the linear analysis.

In total, the statistical analysis in the above data shows that they fit under the Gauss-Markov hypothesis [6]. Overall, the proportion of IDUs changes in the structure of HIV incidence observed between 2006 and 2018, indicating a strong tendency towards its reduction to the overall increase in incidence in the Odessa region.

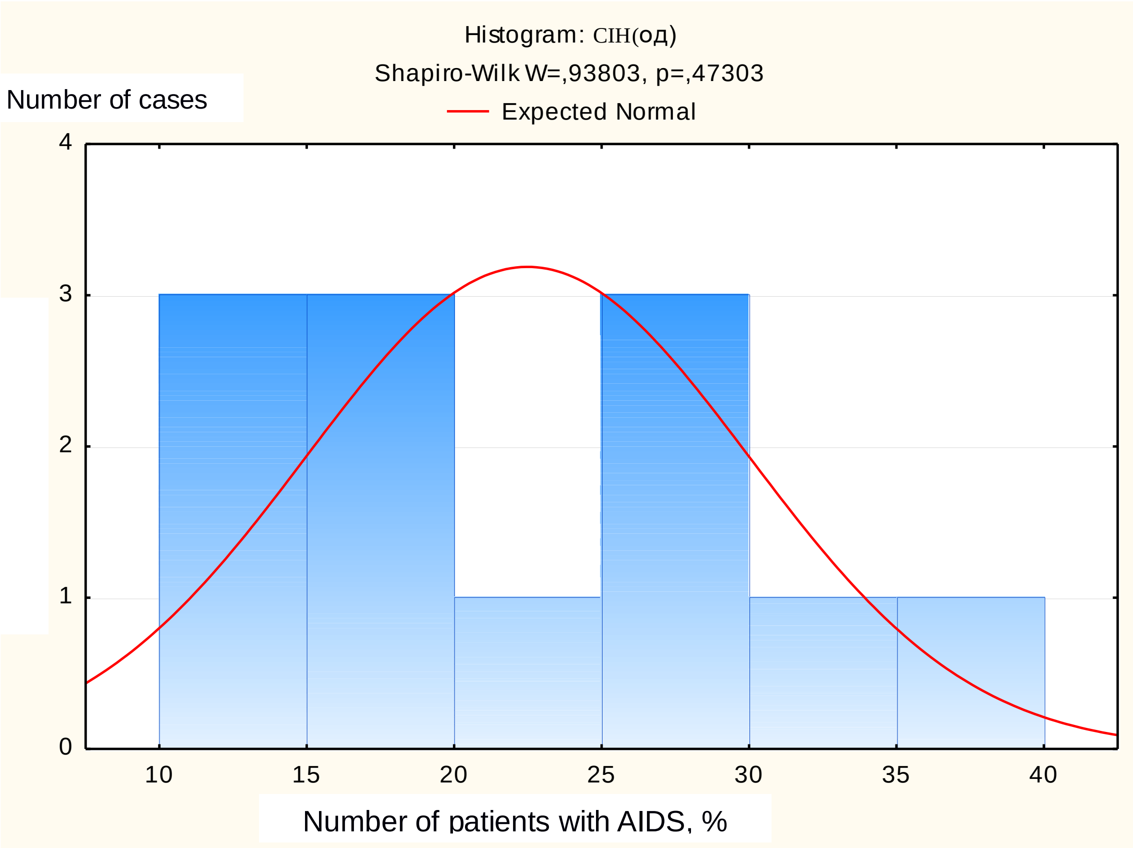

Fig. 1. AIDS incidence distribution (%) among IDU's in the Odessa region

Analysis of the data presented in Fig. 2 shows that for the total number of observed patients' distribution a high coefficient Shapiro-Wilk is at its high significance. However, there is a large variance, though the Shapiro-Wilk criterion is also performed in this case, albeit at a lower value.

To further verify the hypothesis about the normal distribution overall rate of HIV incidence the graph of this distribution balance has been constructed (Figure 3). Because of the fact that the remains are distributed fairly compact concerning the slope of the normal distribution with no emissions, the null hypothesis about the conformity of the normal distribution is accepted.

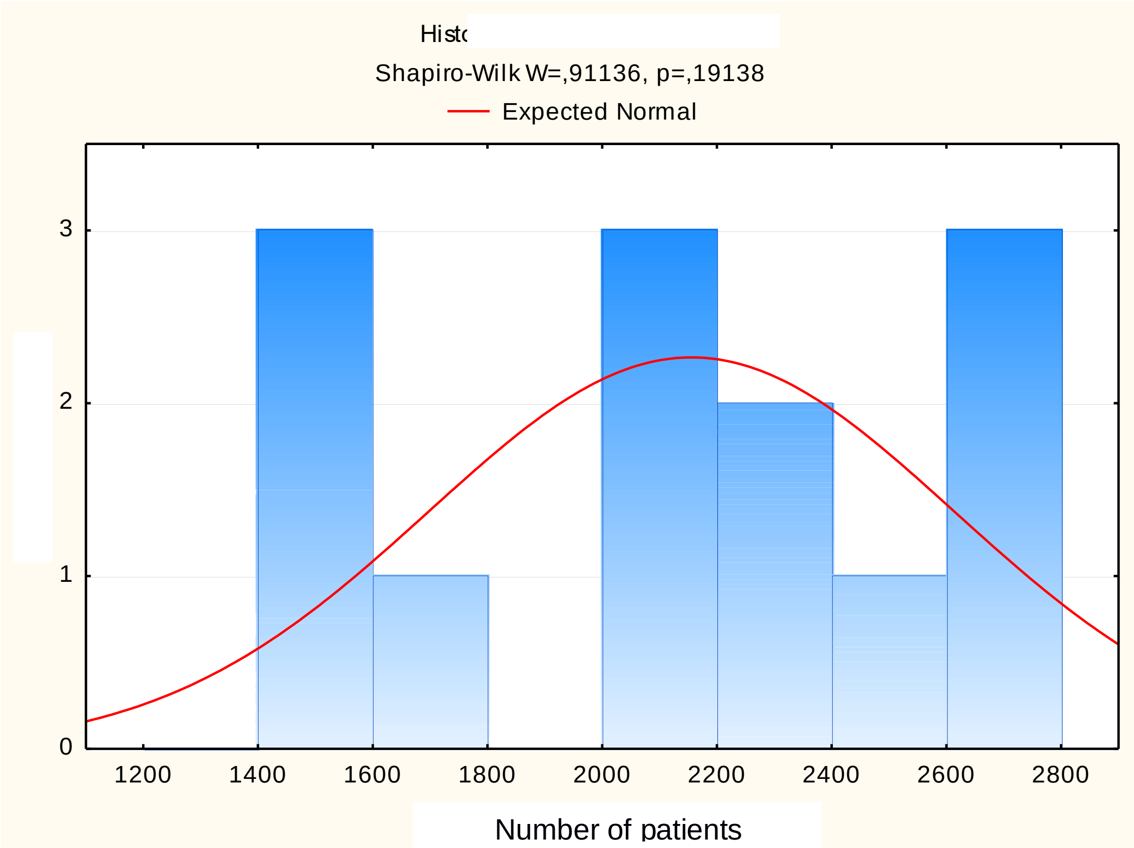

Fig. 2. Number of HIV patients in Odessa region using IDU criteria

In general, the statistical analysis of the data presented in the two samples showed no significant differences between Gauss distribution and the distribution of values in the studied samples. That is, the conducted statistical analysis of the data showed that they fit under the hypothesis of Gauss-Markov.

Similar conclusions were made regarding the presented empirical data on the incidence of HIV in Mykolaiv, Kherson regions and in Ukraine in general.

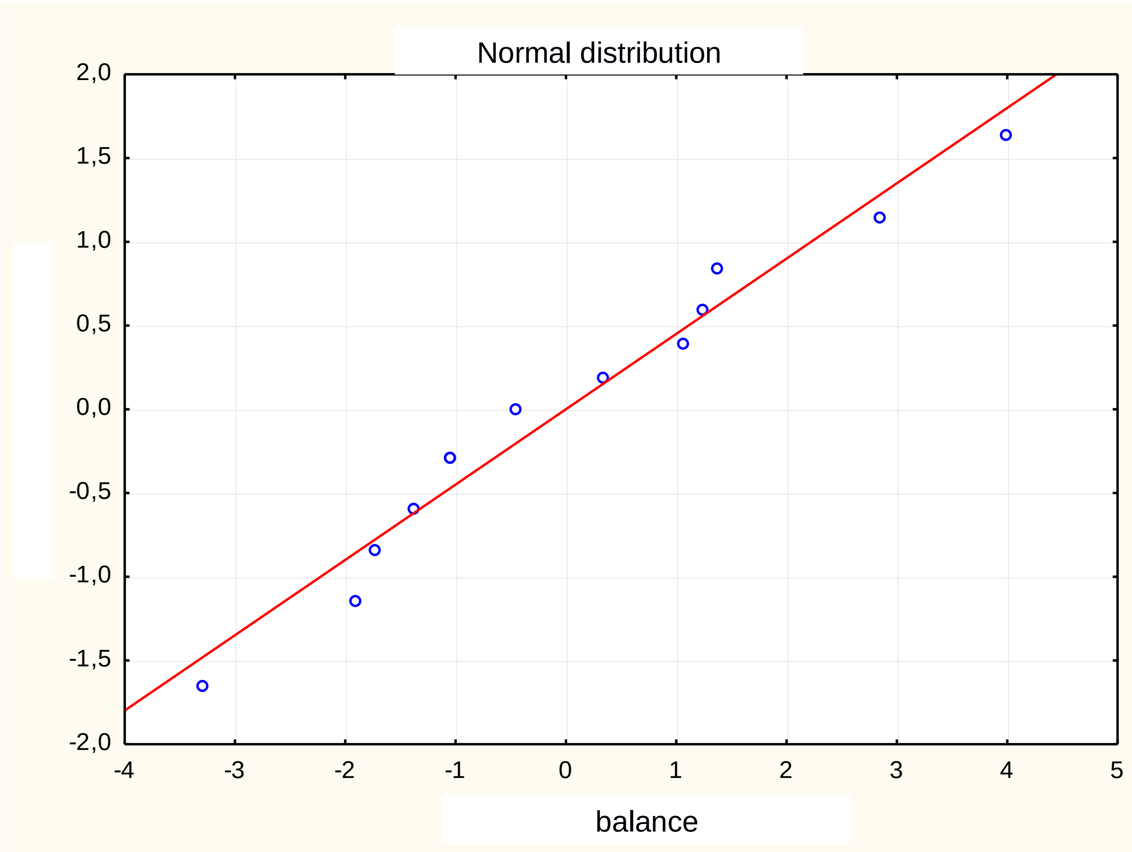

Fig. 3. The normal probability balance graphs for general case of HIV infection in the Odesa region

2. Analysis of the characteristic changes in the distribution depending on the incidence of HIV infection among Odesa, Mykolaiv, Kherson regions and Ukraine (in absolute terms)

A brief analysis of presented empirical data indicates a common trend or a possible relationship between the incidence rates in the regions of the Black Sea region.

In order to determine the presence of such dependence was tested the hypothesis on the relationship between the trends of the HIV infection in Odessa, Mykolaiv and Kherson regions. To do this, we firstly compared the dynamics of morbidity in different regions of the Black Sea region, and subsequently with data on Ukraine throughout the 2006 to 2018 years.

In order to identify the possible statistical relationship between the incidence rate for the accounting period in Odesa and Mykolaiv regions possible correlation between these parameters was analyzed (Figure 4).

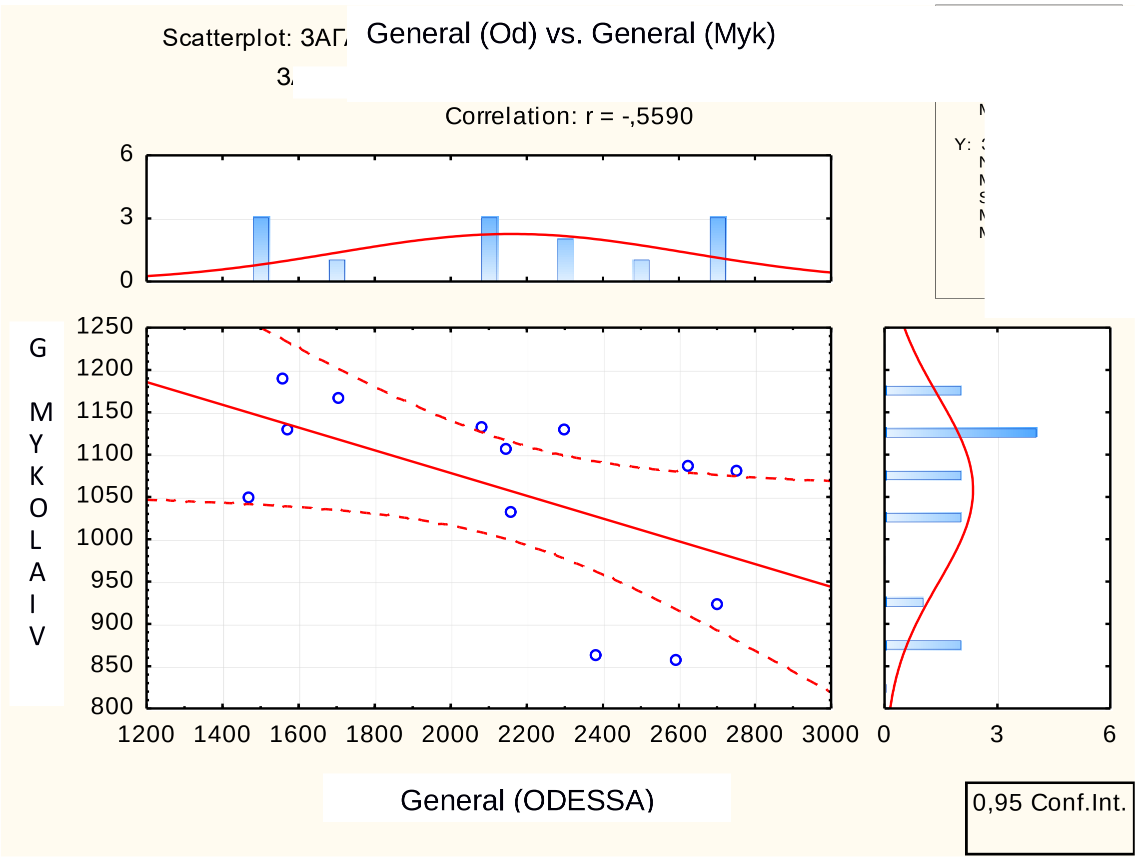

Fig. 4. HIV incidences ratio among the Odesa and Mykolaiv regions (using IDUs absolute criteria).

According to calculations, while comparing morbidity in these areas for the accounting period such a relationship does exist, although the correlation coefficient is negative on average r=-0.56. On the chart, it looks like a continuous downlink though. The dotted line marks on both sides the calculated distribution of the defaults of confidence intervals for confidence level, which is 0.95.

The top (for Odesa region) and right (for Mykolaiv region) in the smaller graphs shows the distribution of incidence quantities allocated in narrow ranges.

For the dependence shown in Fig. 1 it's a specific its decreasing and this trend continues from year to year.

As a quantitative characterization of this phenomenon, we, on the basis of structural and parametric identification process, chose a simple linear mathematical model that makes it possible to compare the incidence of Mykolaiv region with that in the Odessa region. This relationship can be represented as declining linear model in the following form:

(Mykolaiv region) = 1347.1 - 0.134 * INCIDENCE (Odessa region) (Equation 1)

The coefficients of this and all other mathematical models calculated by empirical data and are statistically significant.

Visualization of filtered data and calculation of the proposed formula (see Equation 1) revealed the following trend: during the period subject to analysis, the incidence of HIV in Mykolaiv region decreases, and in the Odessa region, on the contrary – has the tendency to increase.

It is worth noting that the trend of the reduction of the incidence of HIV in Mykolaiv region will likely soon stop because of the so-called "saturation period." This is proven by the change in incidence rates over the past two years. The forecast of the growth of HIV infection in the Odessa region is also disappointing.

Another interesting trend fully characteristic of the Odessa region, namely the increasing lack of correlation between the absolute incidence of HIV and IDU shares in the structure of morbidity.

Fig. 5 shows the average linear correlation between the total absolute indicator of HIV infection and incidence among IDU’s. The dotted line indicated by the upper and lower confidence intervals of 0.95.

Analysis of the dependences presented in Fig. 5 shows the following: according to available statistical dynamics of change of incidence from year to year can say that the absolute increase in the total incidence rate of HIV infection in the Odessa region comes amid the reduction of the incidence of HIV among IDU’s.

Reduction of the rate of HIV incidence among IDU’s may be due to several reasons, including both good performance and results of health care and imperfect accounting system.

The results of the linear statistical analysis of the available data allow us to conclude that this dependence is quite significant, since the resulting correlation coefficient is of high value - $r=0{,}772$, and the character is negative. The data has a compact distribution pattern related to the main slope without significant emissions.

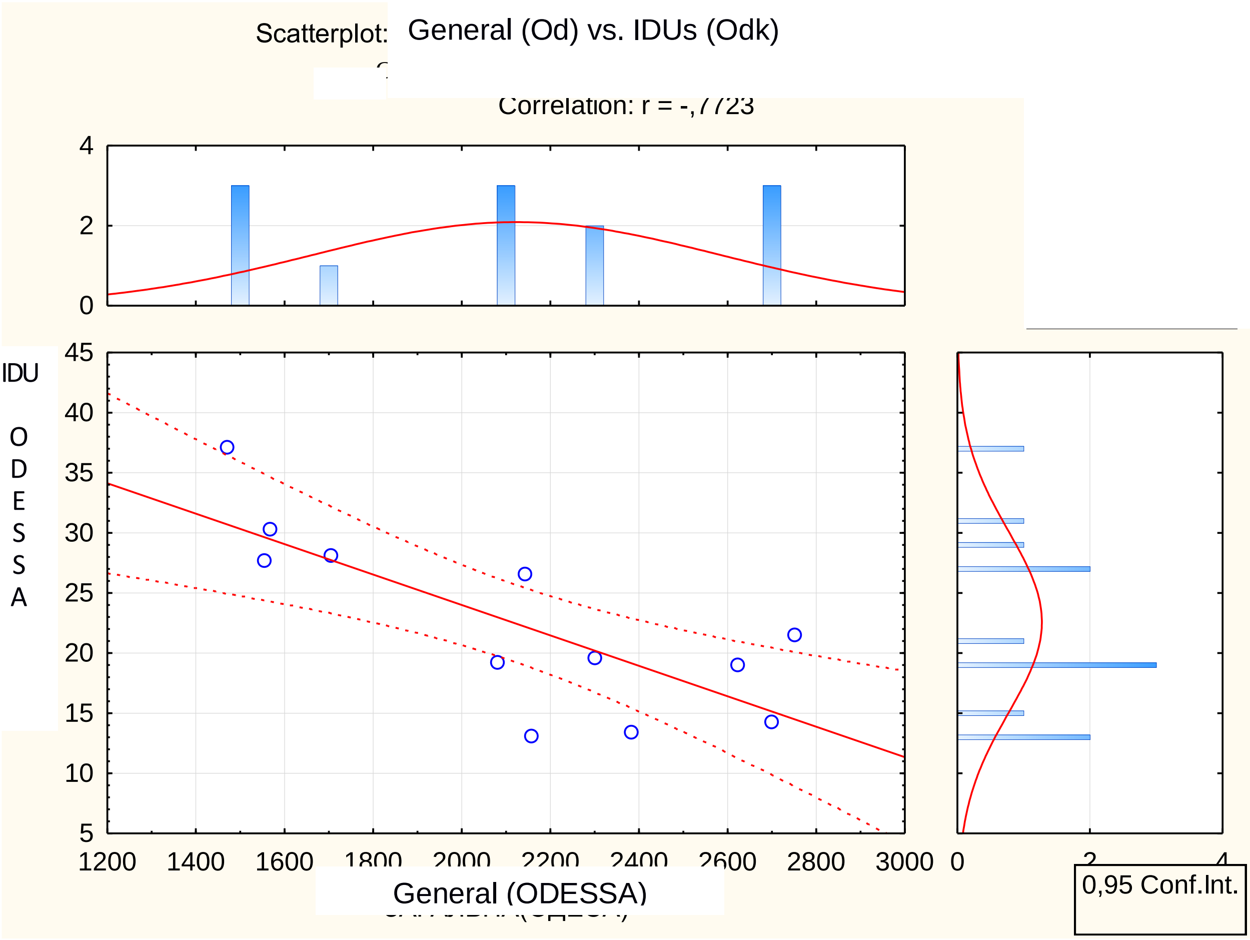

Fig. 5. The ratio between the HIV overall incidence rate and IDU residues in the structure of disease (Odessa region)

This suggests that the obtained dependence is close to reality. The linear equation of this relationship for a given period of time can be represented as follows:

IDU's (Odessa. Reg.) = 49.312 - 0.127 * Total value (Odessa reg.) (Equation 2)

A detailed statistical analysis conducted for each of the regions of the Black Sea region alone, showed some interesting features. Thus, for the Odessa region in annual terms, the spectral analysis [7] the rate of HIV infection among IDU's, the frequency with a period of 4 years was detected (Fig. 6).

However, for the overall incidence of HIV in the Odessa area such an obvious periodicity is absent (Fig. 7). This is possibly due to active migration, especially in the summer, but the data presented in our paper does not have it displayed because no incidence rate monthly distribution was estimated.

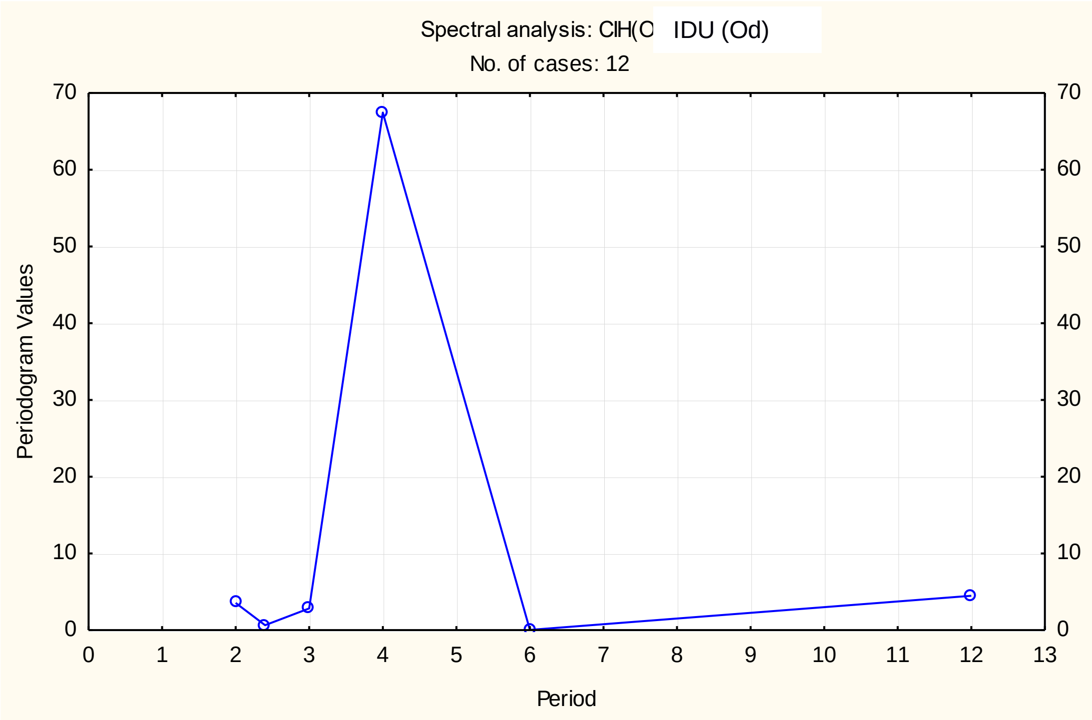

Fig. 6. Periodgraph for the ‘HIV incidence in IDU’ (Odessa region).

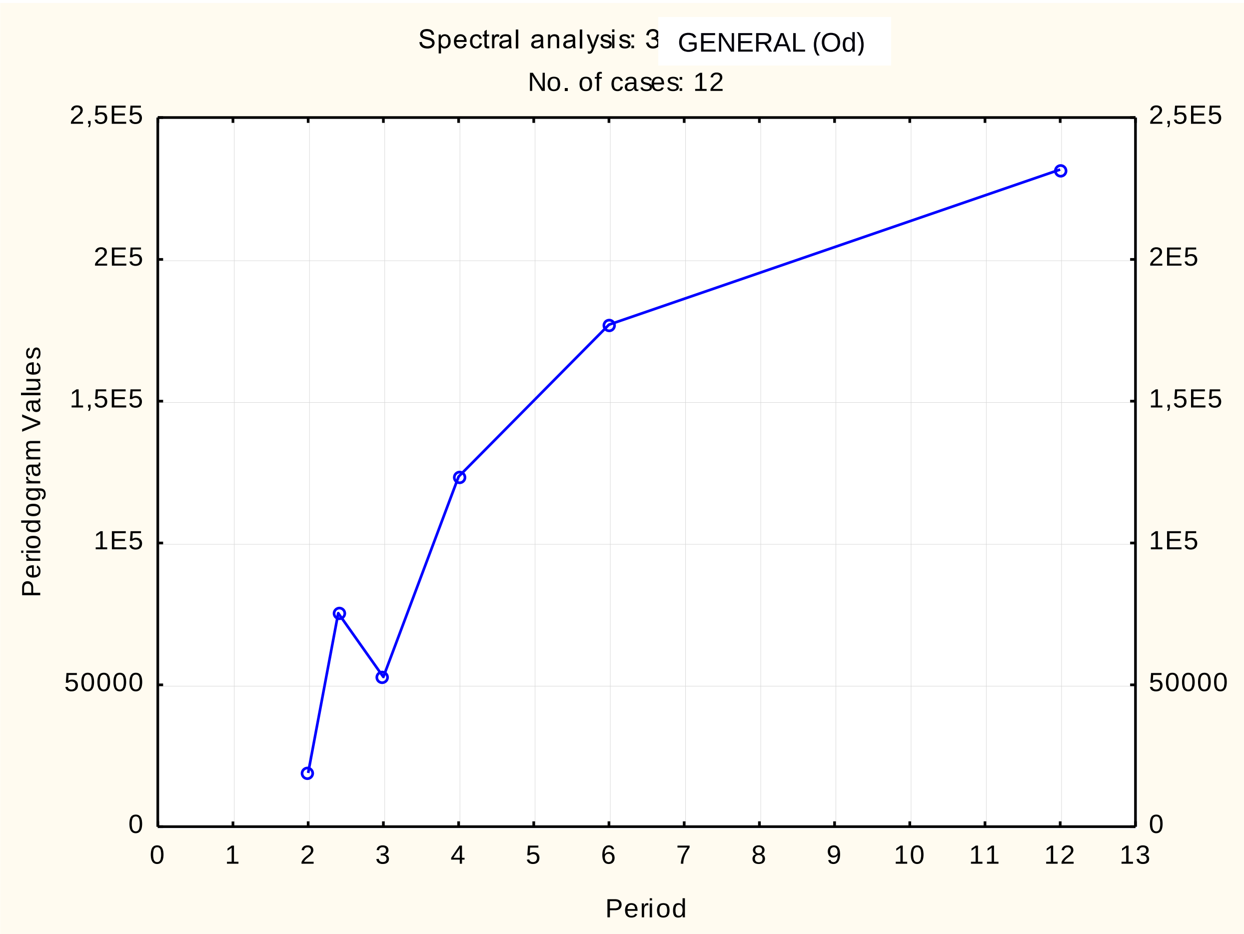

Fig. 7. Periodgraph for the ‘HIV infection general incidence’ criteria (Odessa region)

The chart shown in Fig. 7 manifests a more even distribution of the growing range of general incidence rate (in absolute numbers). In this diagram there is no obvious maximum, the visually empirical curve was also observed. This indicates the lack of annual frequency oscillation of the overall incidence of HIV in the Odessa region rate. A small peak for a period of approximately 2.3 years does not change the overall trend to increase morbidity. This trend shows a steady increase in the future and, if no measures to prevent the spread of infection are taken, it can lead to a significant increase in the incidence as the 5 degree polynomial.

Having done the analysis of the retrospective data on the incidence of HIV in the Odessa area we received a nonlinear relationship, which shows the absolute increase in the incidence as the 5 degree polynomial.

$$\textbf{General morbidity in the Odessa region} = -6 * 10^{16}\ 10^{14} * + 1.5 * t + 1.5 * 10^{11} * t2 + 7.4 * 10^{7} * t^{3} - 1846 * t^{4} + 1.8 * t^{5} \qquad \text{(Equation 3)}$$

If the current trend of the increasing incidence does not change, the proposed extrapolation of data might demonstrate unsatisfactory situation (Fig. 8).

Analysis of the incidence of HIV among IDU’s in Odessa region is different. The data on the incidence of HIV infection from 2006 to 2018 are presented in Fig. 9 and it resembles the character of the curve with saturation by extrapolation, which is advisable to introduce as a second stage functional as follows:

$$\textbf{The incidence of HIV infection (IDU's)} = -3.9 * 10^{7} + 34591,4 * t - 8.6 * t^{2} \qquad \text{(Equation 4)}$$

This dependence implies that while preserving the existing trends in the future the stabilization of incidence of IDU’s is possible at some level.

A slightly different trend when comparing the figures for the overall incidence 2006-2018 in Odessa and Kherson regions (Fig. 10) is seen. Here, according to our research, almost no any relationship between the incidence rates is shown. This difference can be explained by the migration flows differences due to the geographical features of the region.

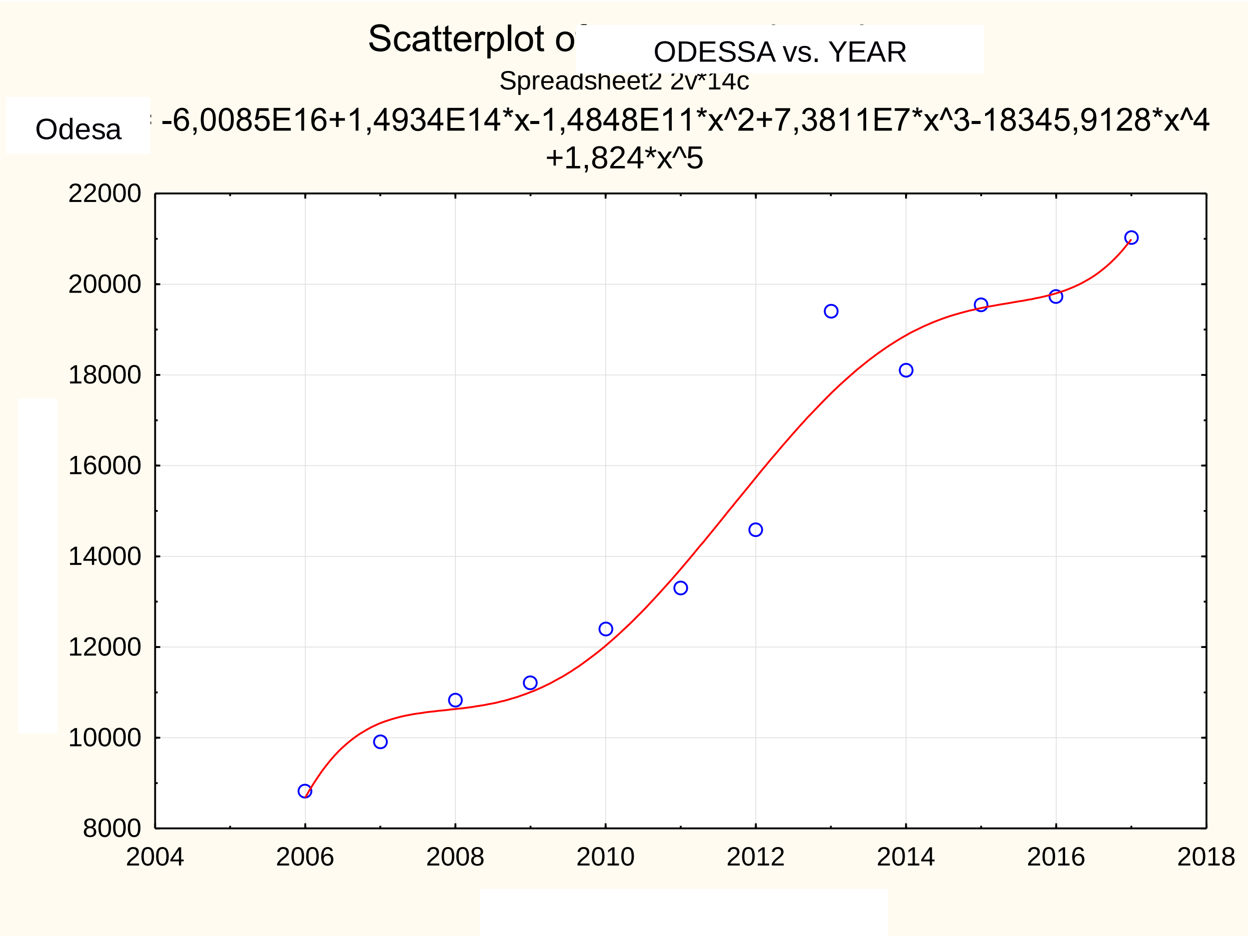

Fig. 8. The graph of the total dependence of HIV incidence in the Odessa region from 2006 to 2018

Analysis of the dependencies (top and side graphics) shows that during the observation no meaningful redistribution of the data has happened, and the nature of the confidence interval did not significantly change.

However, the changed nature of the overall dependence – an obvious correlation between the incidence rates in the Kherson region compared to the Odessa area almost disappeared, leading to a substantial change in the coefficient of correlation. The weak correlation between these values is also proven by a minor correlation coefficient value, which is -0.0134 and, as a result, the lesser value of the slope coefficient in the flowing linear relationship formula:

INCIDENCE (Kherson reg.) = 572.9 - 0.0017 * INCIDENCE (Odessa reg.)

(Equation 5)

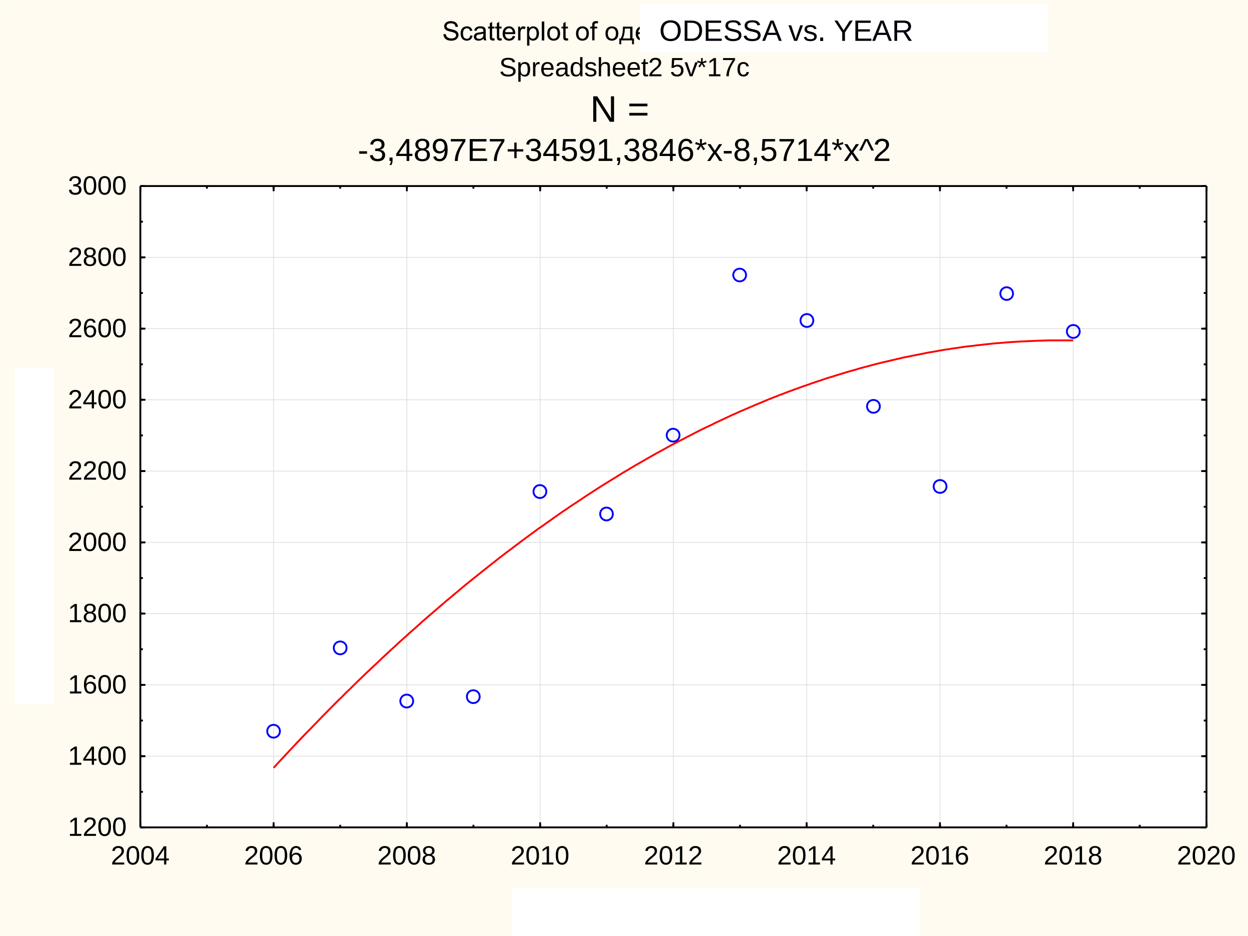

Fig. 9. The incidence of HIV among IDU's.

In Ukraine as a whole, as well as in the Odessa region, during the reporting period the rise of HIV incidence is seen, as evidenced by the data presented in Fig. 11.

Thus was an increase in the coefficient of correlation, it was positive and increased up to p=0.23. In a possible quantitative assessment of this trend in the first approximation to the mathematical model the linear model in the following form can be proposed:

INCIDENCE (Ukr.) = 0.94596 * 16829 + INCIDENCE (Odessa reg.)

(Equation 6)

Moreover, the analysis equation 6 shows the positive trend of the HIV incidence growth in Ukraine as a whole due to the growth of morbidity in the Odessa region, Odessa region is making a significant "contribution" to the overall morbidity in Ukraine. While in Ukraine over the period of the year one could see a steady increase in HIV infection incidence as it is shown in Fig. 11.

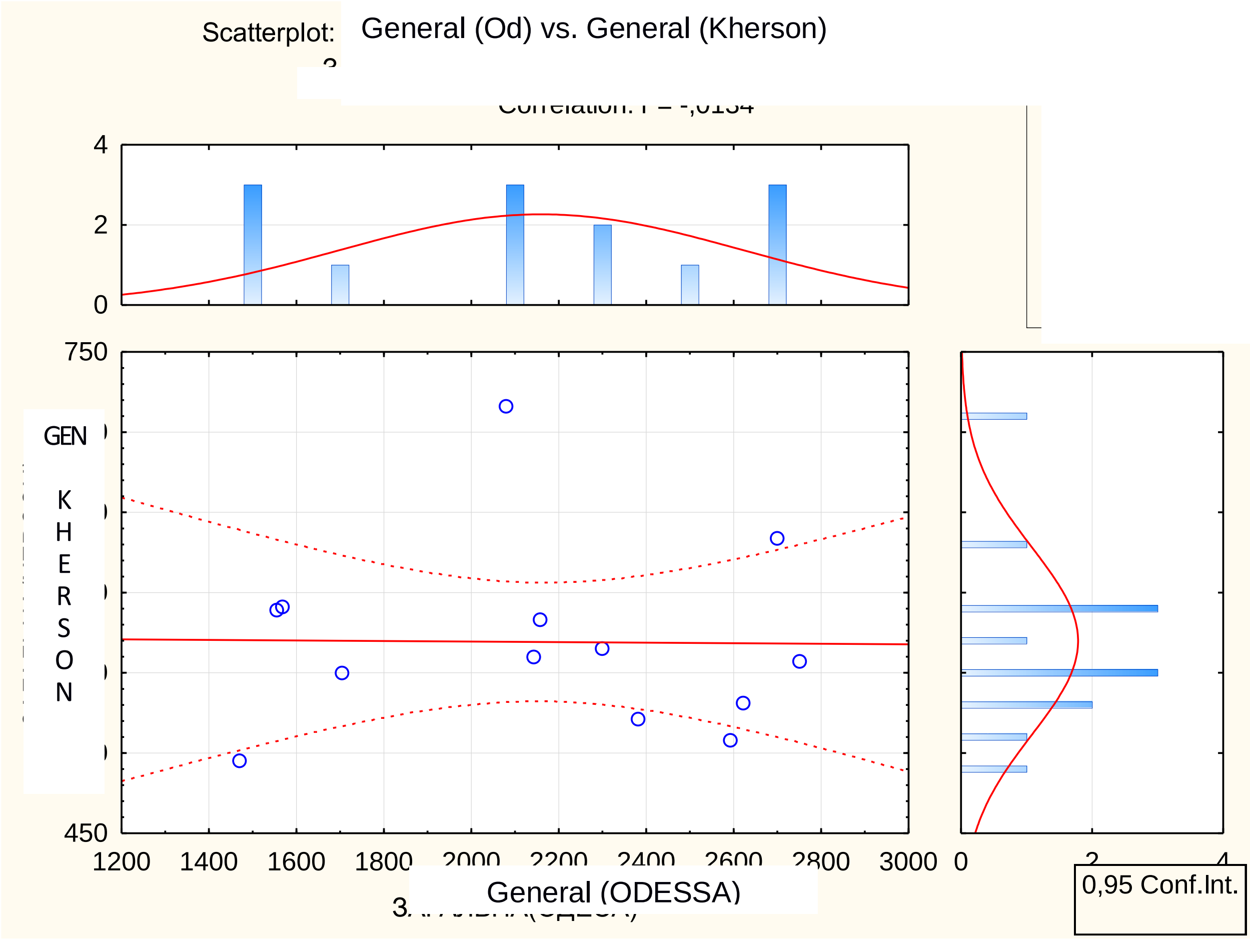

Fig. 10. Dependence of HIV incidences ratio in Odessa and Kherson regions (in absolute numbers).

Conclusions.

1. Odessa and Kherson regions of the Black Sea region demonstrated a steady increase in the HIV incidence; the increase of the incidence makes around 10-20% annually. In the Mykolaiv region, on the contrary, there is an annual reduction in incidence, although this trend weakened in recent years.

2. The nature of fluctuations in the incidence of HIV in the period from 2006 to 2018 differs in Odesa, Mykolaiv, Kherson regions and in Ukraine in general. The correlation between Odesa and Mykolaiv region – is negative, between Odessa and Kherson - almost zero between Ukraine and Odessa region – weakly positive.

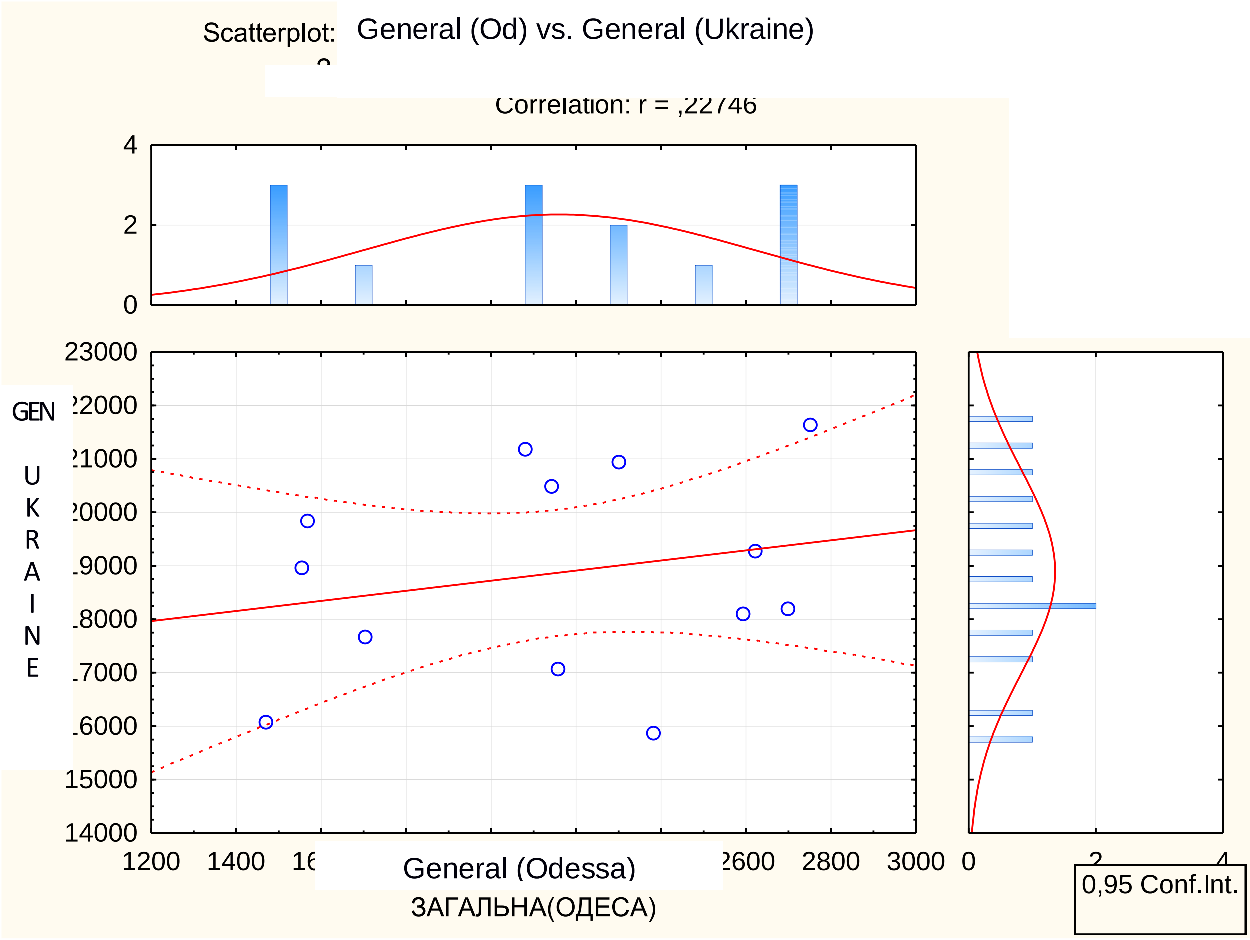

Fig. 11. Dependence of HIV incidences ratio in the Odessa region and totally in Ukraine (in absolute numbers).

REFERENCES

1. Talalayev K.O. Predictive Medicine - Challenges and Achievements. Odessa Medical Journal. 2016; 156 (4) : 69-74 [In Ukrainian].
2. Talalayev K.O. From medical care to predictive medicine. Signs of trend reversal. Journal of Marine Medicine. 2016; 73 (4) : 7-11 [In Ukrainian].
3. Ogloblina O.I., Sushko T.S., Shramko S.V. Elements of number theory: tutorial. - Ministry of Education and Science of Ukraine, Sumy State University. 2015: 185 p [In Ukrainian].
4. Shapiro S.S., Wilk M.B. An analysis of variance test for normality. Biometrika. 1965; 52 (3) : 591-611.
5. Halsey L.G., Curran-Everett D., Vowler S.L., Drummond G.B. The fickle P value generates irreproducible results. Nature Methods. 2015; 12 (3) : 179-185.

6. Stock J.H., Watson M.W. Regression with a Single Regressor: Hypothesis Tests and Confidence Intervals /In: Introduction to Econometrics. The 3rd Ed. Pearson Education Limited, Edinburgh, 2015: 163-164.

7. Kondratiev M.A. Prediction methods and disease spread models. Computer Research and Modeling. 2013; 5 (5): 863–882 [In Russian].