MINISTRY OF EDUCATION AND SCIENCE OF THE REPUBLIC OF KAZAKHSTAN

SULEYMAN DEMIREL UNIVERSITY

UDK______

SULEYMAN DEMIREL UNIVERSITY

ZHUMABEK DILDAR

Analysis and the development of a mathematical model of the children mortality

Specialty: "6M060100 - Mathematics"

Academic Degree: Master of Natural Sciences

"Admitted to defense":

Dean of Faculty
Assist. Prof., PhD Bogdanchikov A.

"___" ____________2019.

Head of Department — Assist. Prof., PhD Borambayeva S.I.

Scientific Advisor — Assist. Prof., PhD Aipenova A.

Kaskelen, 2019

Abstract

In this work constructing of a model equation for the dynamics of children mortality in population of Kazakhstan is described. Mortality is an important aspect of demographic process, closely linked to the socio-economic prospects for the development of society. The mortality rate is also connected with the gender aspect, age structure of countries population, and also in lifespan evaluation. Highlighted and described factors that have the greatest impact on the mortality rate of children of the population of Kazakhstan. The regional peculiarities of this process are considered. In particular, the study has an impact on the lack of qualifications and poor quality of medical personnel and non-compliance with the level of medical service. In addition, another important factor is the dissatisfied state of ecology with man-made causes, the causes of mortality in the context of regions are investigated and complex of factors that influenced the differentiation of mortality among the regions of Kazakhstan are analyzed as well as the dynamics of infant mortality.

Аңдатпа

Бұл жұмыста Қазақстан халқы арасында балалар өлімінің динамикасының модельдік теңдеуін құру және халқы балаларының өлім-жітім деңгейіне барынша әсер ететін факторлар бөлініп сипатталады. Өлім-жітім қоғам дамуының әлеуметтік-экономикалық перспективаларымен тығыз байланысты және демографиялық процестің маңызды аспектісі болып табылады. Өлім-жітім деңгейі гендерлік аспектімен, ел халқының жас құрылымымен, сондай-ақ өмір сүру ұзақтығын бағалаумен байланысты. Бұл үдерістің аймақтық ерекшеліктері карастырылған. Атап айтқанда, медицина қызметкерлерінің біліктілігінің жоқтығы мен сапасының төмендігіне және медициналық қызмет көрсету деңгейінің сақталмауына әсер етеді. Бұдан басқа, тағы бір маңызды фактор техногендік себептерге байланысты экологияның қанағаттанғысыз жай-күйі болып табылады. Өңірлер бөлінісінде өлім-жітімнің себептері зерттеледі және Қазақстан өңірлері арасындағы өлім-жітімнің саралануына әсер еткен факторлар кешені, сондай-ақ нәресте өлім-жітімінің серпіні талданды.

Аннотация

В данной работе описывается построение модельного уравнения динамики детской смертности среди населения Казахстана. Смертность является важным аспектом демографического процесса, тесно связанным с социально-экономическими перспективами развития общества. Уровень смертности также связан с гендерным аспектом, возрастной структурой населения стран, а также с оценкой продолжительности жизни. Выделены и описаны факторы, которые оказывают наибольшее влияние на уровень смертности детей населения Казахстана. Рассмотрены региональные особенности этого процесса. В частности, исследование оказывает влияние на отсутствие квалификации и низкое качество медицинского персонала и несоблюдение уровня медицинского обслуживания. Кроме того, еще одним важным фактором является неудовлетворенное состояние экологии с техногенными причинами. Исследуются причины смертности в разрезе регионов и анализируется комплекс факторов, повлиявших на дифференциацию смертности между регионами Казахстана, а также динамика младенческой смертности.

Contents

1. Introduction

The problem of high mortality among children has been and remains one of the most acute problems in the development of modern Kazakhstan. The mortality rate of children is most adequately reflected in the indicator of life expectancy at birth.

The mortality rate depends on many different factors: the socio-economic development of the country. the environmental situation. the well-being of the population. the level of stress and much more. After fertility. it takes the second place in its importance in the processes of reproduction of the population. has a serious impact on the population size. its structure. and is closely interconnected with all socio-demographic processes. The causes of mortality in Kazakhstan are classified by the main groups: infectious diseases. diseases of the respiratory system. circulatory system, neoplasms. accidents, poisonings and injuries. Mortality of the population is a mirror reflection of the level of socio-economic development of society.

To build a mathematical model in the form of a regression equation. the Ministry of National Economy of the Republic of Kazakhstan used the Statistics Committee. which reflects the mortality rate and the values of its coefficients for the considered causes of death of children in Kazakhstan for 2005–2018. (www.stat.gov.kz). As the investigated factor signs were taken values infant mortality rates (deaths per 100.000 people) for various reasons: from neoplasms. circulatory system diseases. respiratory organs. digestive organs. from some infectious and parasitic diseases, accidental alcohol poisoning. from all types of transport accidents. from drowning. suicides and kills.

1.1 Motivation

Despite the large size of the country's land, given the small number of people living in it, every child who is born is very important to survive in the future of the country. The cause of death of children up to one year are the conditions arising in the perinatal period, inflammation of the respiratory tract, including pneumonia and congenital defects. And the cause of infant mortality is inflammation of the airways, including pneumonia, injuries, burns and acute intestinal infections. According to medical research, at least 70% of infant mortality depends on medicine.

UNICEF identified the most common causes of infant mortality in Kazakhstan: non-compliance with existing standards and algorithms, a shortage of neonatologists, insufficient knowledge and skills, the lack of a unified approach to the use of antibiotics.

In Kazakhstan, over the years, a decline in infant mortality has been observed, largely due to an increase in the well-being of citizens, investments by the government and international organizations in the health care system.

However, children under 5 years of age still die prematurely in the country for reasons that could have been prevented. Most children die in the first year of life (in 2016 - 68.4%). Thus, out of 4.799 children under the age of 5 years who died in Kazakhstan in 2015, 77% (3.719 children) did not live to 1 year.

These indicators are very high and looking I wanted to explore on a mathematical point of view. Make sure these indicators see regression and correlation.

1.2 Aims and Objectives

1. To study the dynamics of infant mortality rates in Kazakhstan (2000-2018 years)
2. Identify factors affecting child mortality
3. To study the mortality of children in the regions of Kazakhstan
4. View people's welfare and ecology where they live and determine how these factors affect the mortality of children
5. Make correlation and regression analysis and write a model. At the end to reveal all the values of interest.

1.3 Thesis Outline

In this dissertation, a mathematical model of infant mortality in the Republic of Kazakhstan, including the South Kazakhstan region, will be created. The thesis work consists of three main sections, nine diagrams, eight tables and twenty-three references are used. The first part of the work begins with an introduction, goals and objectives. The second part sets out the theorems and definitions used in the thesis. In the main part of the work, the types, characteristics of the regression analysis, its advantages in medicine, about the studied scientists were presented. Data for research obtained from the Statistical Committee of the Republic of Kazakhstan, WHO, UNICEF. In the fourth part, the data were analyzed, the necessary materials were identified. In the fifth part of the mathematical model, optimal solutions were developed and made.

2. Preliminaries

Pearson correlation. Pearson's correlation criterion is a method of parametric statistics, which allows to determine the presence or absence of a linear relationship between the two quantitative indicators, as well as to assess its closeness and statistical significance. In other words, the Pearson correlation criterion makes it possible to determine whether there is a linear relationship between changes in the values of two variables. In statistical calculations and conclusions, the correlation coefficient is usually denoted as r_{xy} or R_{xy}.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Gauss-Markov theorem. A steam regression model is considered, in which the observations of Y are related to X with the following dependency: $Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i$. Based on n sample observations, the regression equation is estimated $\hat{Y}_i = \hat{\beta}_i + \hat{\beta}i^{X_i}$. The Gauss — Markov theorem says:

- The data model is correctly specified;
- All X_i are deterministic and not all are equal;
- Errors are not systematic, that is, $E(\varepsilon_i) = 0, \forall i$
- The error variance is the same and equal to some σ^2
- Errors are uncorrelated. $Cov(\varepsilon_i, \varepsilon_j) = 0, \forall i, j$

Student's t-criterion is a generic name for a class of methods for statistical testing of hypotheses (statistical criteria) based on the student's distribution. The most frequent cases of t-test are related to checking the equality of the mean values in two samples.

$$t = \frac{\bar{x} - m}{s_x / \sqrt{n}}$$

The Fisher criterion is used to check the equality of the variances of the two samples. When testing a position hypothesis (hypothesis about the equality of average values in two samples) using Student's criterion, it makes sense to first test the hypothesis about equality of variances.

$$F = \frac{s_1^2}{s_2^2}$$

$$s_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$s_2^2 = \frac{1}{m-1}\sum_{i=1}^{m}(y_i - \bar{y})^2$$

The classical Kolmogorov criterion (sometimes they say Kolmogorov-Smirnov) is intended to test simple hypotheses about the belonging of the analyzed sample to some completely known distribution law.

Let X_n be a sample of independent identically distributed random variables, $F_n(x)$ is an empirical distribution function. F(x) is some "true" distribution function with known parameters. Statistics criterion is determined by the expression:

$$D_n = \sup |F_n(x) - F(x)|$$

Pearson's χ^2 test is a non-parametric method that makes it possible to assess the significance of differences between the actual number of outcomes or qualitative characteristics of a sample falling into each category and the theoretical quantity that can be expected in the studied groups with the validity of the null hypothesis. To put it more simply, the method allows to evaluate the statistical significance of differences between two or more relative indicators.

$$\chi^2 = \sum_{i=1}^{r}\sum_{i=1}^{c}\frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

When i- line number, j- column number.

Bernoulli's distribution in probability theory and mathematical statistics is a discrete probability distribution that models a random experiment of arbitrary nature, with a known probability of success or failure.

Definition: Random variable X has a Bernoulli distribution if it takes only two values: 1 and 0 with probabilities p *and* $q \equiv 1 - q$, respectively. In this way:

$$P(X = 1) = p$$

$$P(X = 0) = q$$

It is customary to say that the event {X = 1} corresponds to "success", and the event {X = 0} - to "failure". These names are conditional, and depending on the specific task can be replaced by the opposite.

3. Literature Reviews

3.1 Child Mortality studies

Mortality as a demographic process is the frequency of death in the social environment. After fertility, it takes the second place in its importance in the processes of reproduction of the population, has a serious impact on the population size, its structure, and is closely interconnected with all socio-demographic processes. The causes of mortality in Kazakhstan are classified by the main groups: infectious diseases, diseases of the respiratory system, circulatory system, neoplasms, accidents, poisonings and injuries. Mortality of the population is a mirror reflection of the level of socio-economic development of society.

At the end of the twentieth century and the beginning of the twenty-first century, significant changes occurred in the state of mortality of the population of Kazakhstan. Back in 1997, in the Address of the President to the people of Kazakhstan "Kazakhstan - 2030. Prosperity, security and improvement of the welfare of all Kazakhstanis" drew attention to the fact that "a strong demographic and migration policy should be put forward as a leading national security priority. If our state bodies continue to be indifferent to this, then we are on the threshold of the 21st century, following Russia in a situation of a "demographic cross," when the population is declining not only due to external migration processes, but also naturally. This trend should be immediately suspended" [1]. Defining the negative features of the demographic development of Kazakhstan in this period, N.A. Nazarbayev pointed to depopulation, "which is dangerous from any point of view. Since 1992, for the first time in 50 post-war years, the population of the republic began to decline" [2].

The overall mortality rate for 1979-1999, increased by 29%, i.e. in 20 years, the number of deaths in Kazakhstan has increased by more than one third. It is possible to single out one of the reasons for the increase in the death rate of the population of Kazakhstan as the demographic aging of the population. The increase in the proportion of the age group over 65 years in 20 years (1979-1999) compared to the previous twenty-year period was almost 10%; environmental degradation in the republic; difficulties of entering the market system of economic relations; commercialization of the health care system and its overall deterioration; the increase in violent mortality associated with the deterioration of the crime situation in the republic.

In the following inter-census period 1999 - 2009. The overall mortality rate remained high, and with all other circumstances contributing to this level of mortality, the aging population played a dominant role (an increase in the composition of the population of older people, among whom the mortality rate is always higher).

It should be noted that the "peak" of mortality occurred in 1995 (168.7 thousand). The mortality rate per 1.000 people thereafter dropped from 10.7 deaths in 1995 to 9.9 in 1999, and then, with slight fluctuations in the direction of increase and decrease, by 2009 reached 9.9 deaths again [3]. By 2015, this figure was 7.57.

Considering the age-specific mortality, it should be especially noted that men of working age were the most affected during this period. In the dynamics of their age-specific mortality, three main trends can be distinguished. In childhood ages (up to 15 years), male mortality has generally declined. Significantly in the age group of 0-5 years. In 1999, this ratio was 5.9, while in 1979 it was 10.9 (by 46%, 1.8 times); in the age group of 5-9 years - by 33.3%, 1.5 times; at the age of 10-14 years - by 14.3% [4].

The second trend was characterized by the fact that from the age range of 15-19 years, male mortality begins to increase gradually (by 13.3%). In the next age group of 20-24 years old, its growth has already reached - 18.2%. And the most significantly deteriorated indicators of age-specific mortality in men 55-59 years of age (by 28.8%), actually by 1/3 This process also affected the younger ages: 25-29 years old (by 15.9%); 30-34 years old (by 14.8%); 35-39 years old (by 7.0%); 40-44 years (10.5%); 45-49 years old (by 9.6%); 50-54 years (by 8.7%). A huge number of men's lives claimed premature death [5].

And a third trend can be found in older ages. In the age group of 60-64 years compared with previous age intervals, the level of male mortality is slightly lower, but in general, this figure increases by 1.2 times during this period [5].

Studying the mortality of men at working age, it should be noted that the situation was close to critical. In the new economic conditions, men were the most vulnerable and susceptible to various external influences. For example, from 1985 to 1993, accidents, homicides, suicides occurred by 25% more. Only in six years, between 1990 and 1996, the death rate separately from homicides increased by 36.2%, from suicides by 35.8% [6].

Consider mortality for the main reasons. Thus, the greatest number of deaths in 1980 occurred from accidents, homicides, suicides and other external influences. In second place were diseases of the circulatory system. In third place is a malignant neoplasm. For 20 years, the structure of causes of mortality in Kazakhstan has changed significantly. In 1999, an increase in mortality was observed in almost all the classes of diseases identified, but diseases of the circulatory organs were already leading. About the same number of lives claimed malignant neoplasms and accidents, murder and suicide, and other external influences. The mortality rate due to malignant neoplasms from 1980 to 1995 significantly increased 1.2 times, then went to a slight decrease. Circulatory diseases in the period from 1980 to 1996, 39% more claimed lives [7]. Then their number began to decline. The same is true for the class of accidents, homicides, suicides and other external influences. Only diseases of the respiratory organs during 19 years as a cause of death decreased from 1980 to 1999 by 48%, 2 times [8, 9, 10]

In the next interdental period, the main causes of death in Kazakhstan remain unchanged. The main causes of death in Kazakhstan for many years remain unchanged. The leading position is occupied by mortality from circulatory system diseases. A high incidence of malignant neoplasms remains. In 2009, the indicator, compared with 2005, decreased by 5.2% (from 192.5 per 100 thousand of the population, to 182.6). At the same time, the advanced forms prevail (16.2%) [11]. Mortality from malignant neoplasms in 2009 occupies the second position in the structure of causes of total mortality, displacing mortality from accidents, poisoning and injuries to third place. Kazakhstan has a high incidence of so-called social diseases, including tuberculosis. In terms of the incidence of tuberculosis, Kazakhstan occupies a leading position among the CIS countries and Europe [11].

The state of mortality is a direct reflection of a significant deterioration in the health of the population of the republic. After the successes achieved in combating infectious diseases with antibiotics, the health care system was unable to effectively resist endogenous diseases - the circulatory system, endocrine system, neoplasms, etc. they have a predominantly endogenous nature, and environmental factors also make a significant contribution to their development — intense rhythm of life, irregular alternation of labor and rest, poor nutrition, unhealthy habits (alcoholism, smoking, etc.), various stressful situations, etc. [12]

Considering the mortality in the regional context, it should be noted the high mortality rate in two regions of the north-eastern and central regions. In 1980, the highest overall mortality rate was recorded in the East Kazakhstan region (10.0). He was just as tall in the Semipalatinsk, Pavlodar, Kokchetau and Kostanay regions. This indicator acquired the least severity in the Mangyshlak region (5.0, which is 2 times less than in the East Kazakhstan region).

In general, during this decade (1979-1999), there has been a tendency for its decline in almost all areas. The worst indicators were also observed in the north-eastern and central regions (in 1999 in the Karaganda region the total mortality rate was 11.6, in the Pavlodar region - 10.5). The lowest mortality rate remained in the South Kazakhstan region (in 1999 - 6.8) and Kyzylorda region (in 1999 - 7.4). The overall mortality rate in the Mangystau region is not high compared to other regions. The same trends continued in the next period.

The regional differences in mortality were also based on a whole range of factors that influenced the differentiation of this process. First of all, these are differences in the age structure of the population (the ratio of the elderly and the old to the young); in objective socio-economic conditions, in the standard of living of the population; in a state of ecological situation; in the number of emigrants, above all, of reproductive age; differences in the level of urbanization of the population; differentiation in the ethnic composition of the regions. In our opinion, differences in the historical conditions of the formation of a particular type of reproduction and differences in behavioral and environmental variables that determine the structure of this process in the regions play a significant role.

The main reasons for the increase in mortality in Kazakhstan were not formed in the last decade of the 20th century, but the socioeconomic crisis significantly accelerated their dynamics. The decline in the standard of living of the population and the lack of adequate compensation led to the accumulation of chronic patients and people in general with poor health. The deterioration of public health will lead to a further decrease in life expectancy over a long period of time. The state of fertility and mortality of the population is a mirror reflection of the processes occurring in society, the standard of living of its population, psychological well-being.

The degree of development of the health care system and the state of health of the population of the Republic objectively reflects infant mortality. In 1979, this figure in Kazakhstan was 32.2; in 1989 - 25.3 (the infant mortality rate for this period decreased by 21.4%), in 1999 it was 20.7. Over the entire period (1979-1999), the coefficient under consideration, according to official statistics, decreased by 35.7%, 1.3 times, which is significant. The reduction in infant mortality, in addition to improving medical care, can be explained by the known effect of reducing the overall birth rate.

Very interesting is the comparison of official data published in the annual statistical reports and the results of the medico-demographic study of Kazakhstan (PMDK), for example, 1999. However, the Kazakhstan Agency for Health Affairs (ARH) and the Kazakhstan Statistics Agency (AS) recorded at that time data on live births and infant mortality, following the criteria established in the former Soviet Union, which differed from the criteria recommended by the World Health Organization (WHO). According to the criteria of the former Soviet Union, a pregnancy that ended with a period of less than 28 weeks was classified as a late miscarriage (even if there were signs of life at the time of delivery). Only in case of survival of a prematurely born child for 7 days, he was considered as a live-born. The outcome of pregnancy, which ended at 28 weeks or more, was classified as a live birth in the presence of respiration, and as stillbirth - in the absence of respiration. In turn, the birth, which ended with the birth of a living child, in the presence of any sign of life (breathing, heartbeat or arbitrary muscle contraction). WHO classified as live birth, regardless of the gestational age at the time of completion of the pregnancy: this also applies to pregnancies that ended with a row of 28 weeks or more [13].

Thus, a number of cases that were classified as cases of live birth or infant mortality, according to the WHO criteria, which were taken as the basis in the PMDK 1999, could be regarded in the official statistics of the Republic of Kazakhstan as late miscarriages or stillbirths [16]. Based on the results of the 1999 medical and demographic study, it was established that over the five-year period preceding the study (1994-1999), the infant mortality rate was 62 per 1.000 births. Over a fifteen-year period of time, this figure dropped from 55 (1984-1989) to 50 (1989-1994). These statistics are evidence of improved living conditions for the population since the late 80s, until the early 90s and deterioration - in the period from the end to the end of the 90s. [13].

Additional evidence that the infant mortality rate has recently increased has been compared with the mortality rates obtained during the PDCM 1995. A 1995 study showed that infant mortality was 40 for the period from 1990 to 1995. The same indicator for the period of 1994-1999, obtained during PMDK 1999, was 62 and reflects an increase of 55% [13].

Official statistics using live birth definitions that do not comply with WHO methodology show that the infant mortality rate in the country after increasing to 28.0 in 1993 began to decline steadily and in 2001 reached 19.4: in 2003 it decreased to 15.3, ranging from 11.8 ‰ in Almaty to 20.8 in Mangistau and 21.8 in Kyzylorda regions [17]. For comparison, we present the infant mortality rates in other countries: Germany - 4.3 ‰, Estonia - 5.7 ‰ (2002), Latvia - 9.8 ‰ (2002), Russia - 13.3 (2002), Uzbekistan - 16, 7 ‰, Ukraine - 10.3 (2002), Kyrgyzstan - 21.2 (2002). However, the medical-demographic studies conducted in Kazakhstan in 1995 and 1999 revealed that the infant mortality rate for 10 years (1989-1999) increased by 24.5%, and the death rate of children under 5 years old - by 26%. In the period from 1989 to 1994, the average infant mortality rate per 1.000 live births was 49.7 and the mortality rate of children under 5 years old was 56.7. [17] From 1995 to 1999, infant mortality per 1,000 live births increased to 61.9, and children under 5 years old - up to 71.4. The medical-demographic study also revealed high rates of neonatal and post neonatal mortality, whose levels increased over 10 years, respectively, by 36.6% and 12.7% [14].

Very significant regional differences in the mortality of children in the first year of life. In areas where there is a high birth rate, a high infant mortality rate was also observed. By 1999, the infant mortality rate remained high for the Mangystau region - 33.2 (exceeded the national average by 38% or 1.6 times), the Pavlodar region - 25.7 (by 19.5%). Over the last inter-census decade (1989-1999), a significant decrease in the number of dead children under 1 year old occurred in the South Kazakhstan region (by 35%) and reached a level of 19.9: Atyrau region by 31% and was set at 21.2. In addition, one can mention Kyzylorda (by 26.2%), Aktobe (by 25.4%), Almaty (by 33%) regions. The infant mortality rate for this period increased in North Kazakhstan (by 12%), Zhambyl (by 20.8%), Pavlodar (by 9.3%) oblasts [14].

Consider the causes of mortality in children. In the first place - respiratory diseases then from conditions arising in the perinatal period (from 28 weeks of gestation, including childbirth and the first 7 days of a child's life) and from congenital anomalies (the infant mortality rate from these causes was 10.5 of 20.0) [15,16].

Consider infant mortality by population structure. In rural areas, it is significantly higher than in urban areas. If in 1979 the infant mortality rate in Kazakhstan was 32.2, then by the structure of the population: in the city - 28.5, and in the village - 35.5. In 1979, the infant mortality rate in rural areas was 20% higher than in urban areas [17]. In 1989, this ratio, respectively, in Kazakhstan -

25.9. in the city - 24.6. in the countryside - 27.3. Over 10 years. the infant mortality rate in rural areas decreased by 23%. but remained above the city by 11% [18]. By 1999. the infant mortality rate in urban areas has decreased by 15.9%: in rural areas - by 30%. The mortality rate of urban children under 1 year in 1999 remained higher than rural by 11% [16].

The high infant mortality rate in rural areas once again raises the problem of the rural health care crisis. Prevention. prevention of pregnancy pathologies in the early stages. primary and. most importantly. qualified health care. unfortunately. significantly reduced for the rural population. Only with serious pathologies can a rural woman receive qualified medical assistance in the city.

The infant mortality rate depends on the quality and timeliness of medical care and the organization of anti-epidemic measures. The adverse situation with infant mortality in the country is also associated with the poor health of women. There has been an increase in the pathologies of pregnancies and childbirth. deterioration of the health of newborns. One of the criteria for a woman's health status and the quality of medical care rendered by her is maternal mortality. the level of which in Kazakhstan decreased by 11% from 1990 to 1999, but remains high compared with economically developed countries. In the CIS countries in 1999. Kazakhstan ranked second after Russia in terms of maternal mortality. The maternal mortality rate in the republic was 49.6 per 100 thousand people. in Russia - 58.2. for comparison in Belarus - 20.4.

In 2000. with the support of UNICEF, an analysis of the causes of infant and child mortality was carried out. which was aimed at finding out not only the causes. but also at what age period. in which regions of the country. and in what types of medical institutions children most often die. The analysis was based on an expert assessment of 3.168 deaths of children who died between the ages of 0 and 5 years from January 1 to December 31. 2000. This accounted for 57.3% of all deaths of children of a specified age in the whole country (5 524). The study presented all five regions of the republic: South. West. East. North and Central. Data collection was carried out in five types of medical institutions: obstetric hospitals. children's somatic and infectious diseases hospitals. central district hospitals. clinics. Data analysis was carried out in the following age categories: early neonatal (168 hours after birth). late neonatal (0-28 days). post-neonatal (28 days - 12 months) and children's periods (1-5 years). The results of this were published in the 2000 Human Development Goals and show that in most cases (62.1%) children died in the early neonatal period, less - in the post-neonatal period (20.5%), whereas in the late neonatal period 8.1%) and from 1 year to 5 years (9.3%) - much less frequently [17].

3.2 Multiple Linear Regression

Linear regression - expressed as a direct dependence of the average value of a value from some other value. Unlike functional dependence $y = f(x)$. when each value of the independent variable x corresponds to one specific value of y. with a linear regression. the same value of x may correspond to different values of y. depending on the case. If. as a result of observation. it is established that for each specific value of x there are some (n) values of the variable y. then the dependence of the arithmetic mean values of y on x is a regression in the statistical sense.

If the established dependence can be written as a straight line equation

$y = ax + b$. then this regression dependence is called linear regression. Multiple linear regression is said to be established when a relationship is established between two variables (x and y). Multiple linear regression is also called single-factor linear regression. since one factor (independent variable x) affects the resulting variable (dependent variable y). To understand linear regression. we have to check the correlation.

Correlation dependence is a probabilistic dependence between quantities. which occurs when one of the variables depends not only on this second. but also on a number of random factors. or when among the conditions on which both are dependent. there are common conditions for both of them. That is. the correlation dependence differs from the functional dependence. in which one value depends only on the second and there is a one-to-one correspondence: the value of one value corresponds to a strictly defined value of the second value. Therefore, although with the correlation dependence, the results of observation are located at some approximation to a straight line, they do not lie on a straight line, but only approach it. The tightness of linear dependence characterizes the coefficient of pairwise linear correlation (1). For a more substantive presentation. it should be noted that here we are talking about the Pearson pair-correlation coefficient. There are other types of correlation coefficients. for example, the Spearman correlation coefficient. the Kendall correlation coefficient. and others. But the Pearson correlation coefficient is used in most cases. since most often it is assumed that the distribution of variables is normal or insignificantly different from normal. Such a distribution is a condition for the use of the Pearson correlation coefficient.

Correlation analysis is a set of correlation-based methods for detecting correlation dependence between random variables or signs based on the theory of correlation. The correlation field and the correlation table are ancillary tools in the analysis of sample data. When applied to the coordinate plane of two-dimensional sample points receive a correlation field. By the nature of the location of the field points. you can make a preliminary opinion about the form of dependence of random variables (for example. that one quantity increases or decreases on average with another). For numerical processing. the results are usually grouped and presented in the form of a correlation table. In each cell of the correlation table, the numbers of those pairs whose components fall into the corresponding intervals of the grouping for each variable are given. Assuming the lengths of the intervals of the group are equal among themselves. they choose the centers of the intervals and the number of pairs as the basis for calculations. When correlation analysis is usually not indicating which of the factors is dependent. and which - independent. Also. the task of the correlation analysis does not include setting up a form of dependence between variables and. accordingly, drawing up a formula reflecting the form of dependence. This is part of regression analysis tasks.

The value characterizing the direction and strength of the relationship between by signs.

The correlation coefficient. which by one number gives an idea of the direction and strength of the relationship between the signs. the limits of its fluctuations from 0 to ± 1.

Strength of correlation

strong: ± 0.7 to ± 1

medium: ± 0.3 to ± 0.699

weak: 0 to ± 0.299

Any economic, sociological indicator is most often influenced not by one, but by several factors. For example, the demand for some good is determined not only by the price of this good, but also by price points for substitute and complementary goods, consumer income and many other factors. In this case, instead of pair regression, multiple regression is considered.

$$\hat{y} = f(x_1, x_2, ..., x_p) \qquad (1)$$

Multiple regression analysis is the development of a pair of regression analysis in cases where the dependent variable is associated with more than one independent variable. Much of the analysis is an immediate extension of the paired regression model, but here too there are some new problems, two of which should be highlighted. The first problem concerns the study of the influence of a particular independent variable on a dependent variable, as well as the distinction between its effect and the effects of other independent variables. The second important problem is the specification of the model, which consists in answering the question of which factors should be included in the regression (1) and which should be excluded from it. In the sequel, the presentation of the general questions of multiple regression analysis will be carried out, delimiting these problems. Therefore, we first assume that the specification of the model is correct. The most used and simplest of the multiple regression models is the linear model of multiple regression:

$$y = \alpha' + \beta_1' x_1 + \beta_2' x_2 + ... + \beta_p' x_p + \varepsilon \qquad (2)$$

By the mathematical meaning, the β'_j coefficients in equation (2) are equal to the partial derivatives of the resultant attribute y for the corresponding factors:

$$\beta_1' = \frac{\partial y}{\partial x_1},\ \beta_2' = \frac{\partial y}{\partial x_2},\ ...,\ \beta_p' = \frac{\partial y}{\partial x_p} \qquad (3)$$

The parameter α is called the free term and determines the value of y in the case when all explanatory variables are zero. However, as in the case of pair regression, factors in their economic content often cannot take null values, and the value of a free member does not have an economic sense. At the same time, in contrast to the pair regression, the value of each regression coefficient β'_j is equal to the average change in y with increasing x_j by one unit only under the condition that all other factors remain unchanged. The value of ε is a random error of regression dependence.

In passing, we note that it is most simple to determine the β'_j parameter estimates by changing only one factor x_j, while keeping the values of other factors unchanged. Then the task of estimating parameters would be reduced to a sequence of tasks of a pair of regression analysis for each factor. However, such an approach, widely used in natural science (physical, chemical, biological), in economics is not acceptable. The economist, in contrast to the experimenter - naturalist, is deprived of the opportunity to regulate certain factors, since it is not possible to ensure the equality of all other conditions for assessing the influence of one factor under study.

Obtaining estimates of the parameters $\alpha', \beta_1', \beta_2' ..., \beta_p'$ of the regression equation (2) is one of the most important tasks of multiple regression analysis. The most common method of solving this problem is the method of least squares (MLS). Its essence is in minimizing the sum of the squares of the deviations of the observed values of the dependent variable y from $\hat{y}$ its values obtained using the regression equation. Since the $\alpha', \beta_1', \beta_2' ..., \beta_p'$ parameters are random variables. it is impossible to determine their true values from the sample. Therefore. instead of the theoretical regression equation (2). the so-called empirical regression equation is estimated. which can be represented in the form:

$$y = a + b_1 x_1 + b_2 x_2 + ... + b_p x_p + e \qquad (4)$$

$a, b_1, b_2, ..., b_p$ - estimates of theoretical values or empirical regression coefficients. e is the deviation estimate of ε. Then the calculated expression has the form:

$$\hat{y} = a + b_1 x_1 + b_2 x_2 + ... + b_p x_p \qquad (5)$$

Let there be n observations of explanatory variables and the corresponding values of the effective feature:

$$(x_{i1}, x_{i2}, ..., x_{ip}, y_i), \quad i = \overline{1, n} \qquad (6)$$

To uniquely determine the values of the parameters of equation (4). the sample size n must not be less than the number of parameters. $n \geq p + 1$. Otherwise. the parameter values cannot be determined unambiguously. If $n=p+1$, parameter estimates are calculated uniquely without MLS by simply substituting the values (6) into expression (5). It turns out a system of (p + 1) equations with the same number of unknowns. which is solved by any method applied to systems of linear algebraic equations. However. from the point of view of the statistical approach. this solution to the problem is unreliable, since the measured values of the variables (6) contain various types of errors. Therefore. to obtain reliable estimates of the parameters of equation (5), the sample size must significantly exceed the number of parameters determined from it. Practically, as mentioned earlier. the sample size must exceed the number of parameters at x_i in equation (4) by 6-7 times.

To conduct an analysis within the framework of a linear model of multiple regression, it is necessary to fulfill a number of prerequisites of MLS. Basically. these are the same prerequisites as for paired regression. but here you need to add assumptions specific to multiple regression:

- The specification of the model is (2).
- Lack of multicollinearity: there is no strict linear relationship between explanatory variables. which plays an important role in the selection of factors in solving the problem of model specification.
- Errors $\varepsilon_i, i = \overline{1, n}$ have a normal distribution $(\varepsilon_i \sim N(0, \sigma))$.

The feasibility of this condition is needed to test statistical hypotheses and construct interval estimates. With the fulfillment of all these prerequisites. a multidimensional analogue of the Gauss - Markov theorem holds.

Quality Analysis of the Empirical Multiple Equation linear regression.

Checking the statistical quality of the estimated regression equation is carried out, on the one hand, by the statistical significance of the parameters of the equation, and on the other hand, by the overall quality of the regression equation. In addition, the feasibility of the MLS prerequisites is checked.

At first we will consider the first two types of checks and the questions connected with them. Some prerequisites of MLS and checks of their feasibility will be considered separately.

As in the case of pair regression, the statistical significance of the parameters of multiple linear regression with p factors is checked on the basis of t - statistics:

$$t_{b_j} = \frac{b_j}{m_{b_j}} \quad \left(\text{or } t_a = \frac{a}{m_a} \right) \qquad (7)$$

When $m_{b_j}\left(m_a\right)$ - called standard error parameter's $b_j\left(a\right)$. It is defined as follows. Denote the matrix:

$$Z^{-1} = \left(X'X\right)^{-1},$$

and in this matrix we denote j diagonal element like this z_{jj}'. Then the sample variance of the empirical regression parameter is:

$$m_{b_j}^2 = s^2 z_{jj}', \quad j = \overline{1, p} \qquad (8)$$

and for the free term the expression has the form:

$$m_a^2 = s^2 z_{00}', \qquad (8')$$

if we assume that in the matrix Z^{-1} indices vary from 0 to p. s^2 - unbiased estimate of the variance of the random error ε:

$$s^2 = \frac{\sum e_i^2}{n - p - 1} \qquad (9)$$

Standard errors of regression parameters are:

$$m_{b_j} = \sqrt{m_{b_j}^2} \left(\text{или} \quad m_a = \sqrt{m_a^2} \right) \qquad (10)$$

The derived by expression (6) t - statistics for the corresponding parameter has a Student's distribution with a number of degrees of freedom (n-p-1). At the required level of significance α, this statistic is compared with the critical point of the Student's $t(\alpha; n-p-1)$ distribution (two-sided).

If $|t| > t(\alpha; n-p-1)$, then the corresponding parameter is considered statistically significant, and zero is a hypothesis in the form of $H_0: b_j=0$ or $H_0: a=0$ is rejected.

Otherwise ($|t| < t(\alpha; n-p-1)$) the parameter is considered statistically insignificant, and zero — the hypothesis cannot be rejected. Since b_j does not differ significantly from zero, the factor x_j is not linearly related to the result. Its presence among explanatory variables is not justified from a statistical point of view. Without having any serious effect on the dependent variable, it only distorts the real picture of the relationship. Therefore, after establishing the fact that the coefficient b_j is statistically insignificant, it is recommended to exclude the variable x_j from the regression equation. This will not lead to a significant loss of quality of the model, but will make it more specific.

A strict test of the significance of the parameters can be replaced by a simple comparative analysis.

If $|t| \le 1$ those $b_j < m_{b_j}$, the coefficient is statistically insignificant.

If $1 < |t| \le 2$ those $b_j < 2m_{b_j}$, the coefficient is relatively significant. In this case, it is recommended to use the table of Student's critical distribution points.

If $2 < |t| \le 3$, then the coefficient is significant. This statement is guaranteed for $(n-p-1) > 20$ and $\alpha \ge 0{,}05$.

If $|t| > 3$, then the coefficient is considered strongly significant. The probability of error in this case with a sufficient number of observations does not exceed 0.001.

The analysis of the significance of the coefficient b_j can be approached in another way. For this purpose, an interval estimate of the corresponding coefficient is constructed. If we set the significance level α, then the confidence interval, in which the unknown value of the parameter falls with probability $(1-\alpha)$, is determined by the inequality:

$$b_j - t(\alpha; n-p-1) \cdot m_{b_j} < \beta_j' < b_j + t(\alpha; n-p-1) \cdot m_{b_j} \quad (11)$$

Or

$$b_j - t(\alpha; n-p-1) \cdot m_{b_j} < \beta_j' < b_j + t(\alpha; n-p-1) \cdot m_{b_j} \quad (11')$$

If the confidence interval does not contain a zero value, then the corresponding parameter is statistically significant, otherwise the hypothesis of a zero value of the parameter cannot be rejected.

To check the overall quality of the regression equation, the coefficient of determination R^2 is used, which is generally calculated by the formula:

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (y_i - \bar{y})^2} \quad (12)$$

It shows, as in the pair regression, the proportion of the total variance y, explained by the regression equation. Its values are between zero and one. The closer this coefficient is to unity, the more the regression equation explains the behavior of y.

For multiple regression, R^2 is a non-decreasing function of the number of explanatory variables. Adding a new explanatory variable never reduces the value of R^2. Indeed, each following explanatory variable can only supplement, but not reduce, the information explaining the behavior of the dependent variable.

In formula (11), the residual dispersion is used, which has a systematic error in the direction of decreasing, the more significant, the more parameters are determined in the equation of regression for a given observation volume n. If the number of parameters $(p + 1)$ approaches n, then the residual dispersion will be close to zero and the coefficient of determination will approach unity, even if the factors are weakly related to the result. Therefore, in the numerator and denominator of the fraction in (25) a correction is made for the number of degrees of freedom of the residual and total variance, respectively:

$$\overline{R}^2 = 1 - \frac{\sum e_i^2 /(n-p-1)}{\sum (y_i - \overline{y})^2 /(n-1)} \qquad (13)$$

Since the value of (11), as a rule, increases with the addition of the explanatory variable to the regression equation even without sufficient grounds, the adjusted coefficient (26) compensates for this increase by imposing a "penalty" for increasing the number of independent variables. Rewrite (12) in the following way:

$$\overline{R}^2 = 1 - \left(1 - R^2\right)\frac{n-1}{n-p-1} = \frac{n-1}{n-p-1}R^2 - \frac{p}{n-p-1} = R^2 - \frac{p}{n-p-1}\left(1 - R^2\right) \qquad (14)$$

As p increases, the $p / (n-p-1)$ ratio increases and, consequently, the size of the adjustment of the R^2 coefficient increases.

From (27) it is $\overline{R}^2 < R^2$ obvious that for $p > 1$. As p grows, $\overline{R}^2$ it grows slower than R^2. In other words, it is corrected in the direction of decrease with increasing number of explanatory variables. However $\overline{R}^2 = R^2$, only when $R^2 = 1$. $\overline{R}^2$ it may even take negative values (for example, at $R^2 = 0$). Therefore, for the correction of (26) there is no rigorous mathematical justification.

$\overline{R}^2$ is proved that increases with the addition of a new explanatory variable if and only if t - statistics for this variable modulo more than one. It does not follow from this, as one would assume that an increase means $\overline{R}^2$ an improvement in the specification of the equation. Nevertheless, the addition of new factors to the model takes place as long as the corrected coefficient of determination increases.

Usually, data are presented for both R^2 and $\overline{R}^2$, which are summary measures of the overall quality of the regression equation. However, one should not absolutize the significance of the determination coefficients. There are many examples of incorrectly constructed models with high coefficients of determination. Therefore, the coefficient of determination is currently considered only as one of a number of indicators that need to be analyzed in order to clarify the model under construction.

Analysis of the statistical significance of the coefficient of determination is carried out on the basis of testing zero - hypothesis H_0: $R^2=0$ versus alternative hypothesis H_1: $R^2>0$. To test this hypothesis, use the following F - statistics:

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-p-1}{p} \qquad (15)$$

The value of F for the fulfillment of the MLS prerequisites and for the validity of zero — the hypothesis has a Fisher distribution. From (28) it can be seen that the indices F and R^2 are equal or not equal to zero simultaneously. If $F = 0$, then $R^2 = 0$, and the regression line $Y = \overline{Y}$ is best in MLS, and, therefore, the value of y is linearly independent of $x_1, x_2, ..., x_p$.

To test the zero-hypothesis for a given level of significance α on the tables of critical points of the Fisher distribution, the critical value $F_{table}(\alpha;\ p;\ n\text{-}p\text{-}1)$ is found. If $F>F_{table}$, zero - the hypothesis is rejected, which is equivalent to the statistical significance of R^2, i.e. $R^2> 1$.

An equivalent analysis can be proposed by considering another zero - hypothesis that is formulated as $H_0 : \beta_1'= \beta_2'= ... = \beta_p'= 0$. This hypothesis can be called a hypothesis about the overall significance of the regression equation. If this hypothesis is not rejected, then the conclusion is made that the cumulative effect of all p explanatory variables $x_1, x_2, ..., x_p$ and a on the dependent variable y can be considered statistically insignificant, and the overall quality of the regression equation is low.

Testing of such a hypothesis is carried out on the basis of a dispersive analysis of the comparison of the explained and residual dispersions, i.e. zero - hypothesis is formulated as H_0:$D_{fact}=D_{resid}$ versus the alternative hypothesis H_1:$D_{fact}>D_{resid}$. In this case, F - statistics is built:

$$F = \frac{\sum(\hat{y}_i - \overline{y})^2 / p}{\sum(y_i - \hat{y}_i)^2 / (n-p-1)} \qquad (16)$$

Here in the numerator is the explained (factor) variance per one degree of freedom (the number of degrees of freedom is equal to the number of factors, i.e. p). In the denominator - residual dispersion for one degree of freedom. Its number of degrees of freedom is $(n\text{-}p\text{-}1)$. The loss $(p + 1)$ of the degree of freedom is connected with the necessity to solve the system $(p + 1)$ of linear equations when determining the parameters of the empirical regression equation. If we take into account that the number of degrees of freedom of the total variance is $(n\text{-}1)$, then the number of degrees of freedom of the explained dispersion is equal to the difference $(n\text{-}1)$ - $(n\text{-}p\text{-}1)$, i.e. p. It should be noted that expression (16) is equivalent to (15).

An analysis of F statistics allows us to conclude that in order to accept the hypothesis that all linear regression coefficients are equal to zero simultaneously, the coefficient of determination R^2 should differ significantly from zero. Its critical value decreases with an increase in the number of observations and can become arbitrarily small.

All previous arguments and conclusions concerning classical multiple regression were based on the assumption that we are dealing with the correct specification of the model. In this case, the model specification refers to the choice of explanatory variables. In this regard, it is important to consider two issues that make sense in multiple regression, when a researcher deals with several factors: possible multicollinearity of factors and partial correlation. The latter is particularly closely related to the procedures for the step-by-step selection of variables.

The inclusion of a set of factors into the multiple regression equation is primarily due to the researcher's idea of the nature of the relationship of the modeled indicator with other economic phenomena. The factors included in the multiple regression must meet the following requirements:

- They must be quantifiable. If it is necessary to include a qualitative factor in the model that does not have a quantitative measurement, then it needs to be quantified. For example, in a yield model, soil quality is set as points: in the value model of non-movable objects, the location of the real estate is taken into account: areas can be ranked.
- Factors should not be correlated with each other and even more so in an exact functional relationship.

The inclusion in the model of factors with high cross-correlation, when, for example, $r_{yx_1} < r_{x_1x_2}$ for addiction $y = a + b_1x_1 + b_2x_2 + e$ can lead to undesirable consequences — the system of normal equations can be ill-conditioned and entail the instability and unreliability of estimates of the regression coefficients.

If there is a high correlation between the factors, then it is impossible to determine their isolated effect on the resultant indicator and the parameters of the regression equation turn out to be non-interpreted. Thus, in the above dependence with two factors, it is assumed that the factors x_1 and x_2 are independent of each other, i.e. $r_{x_1x_2} = 0$. Then we can say that the parameter b_1 measures the strength of the influence of the factor x_1 on the result y with a constant value of the factor x_2. If $r_{x_1x_2} = 1$, then with a change in the factor x_1, the factor x_2 cannot remain unchanged. Hence, b_1 and b_2 cannot be interpreted as an indicator of the separate influence of x_1 and x_2 on y.

As mentioned earlier, the addition of a new factor in the regression leads to an increase in the coefficient of determination and a decrease in the residual variance. However, these changes may be insignificant, and not every factor should be introduced into the model.

3.3 Application of Linear Regression on population-related studies

Statistics in medicine is one of the tools for analyzing experimental data and clinical observers. However, this is not the only task of statistics in medicine. The mathematical apparatus is widely used for diagnostic purposes, in solving classification problems and in the search for new laws, for the formulation of new scientific hypotheses. The use of statistical programs involves knowledge of the basic methods and stages of statistical analysis: their requirements, necessity and sufficiency. This is based on the principle of the detailed presentation of the formulas that make up statistical methods, as well as their essence and rules of application.

Statistical processing of medical research is based on the principle that what is true for a random sample is also true for the general population (population) from which this sample was obtained. However, choosing or recruiting a truly random sample from the general population is practically very difficult. Therefore, one should strive to ensure that the sample is representative in relation to the studied population, i.e. sufficiently adequately reflecting all possible aspects of the studied condition or disease in population, which contributes to a clear formulation of goals and strict adherence to the criteria for inclusion and exclusion, both in the study and in statistical analysis.

Statistical data can be presented as quantitative (numerical continuous or discrete), and qualitative (categorical ordinal or nominal) variables. Required to clearly indicate the type (type) of the variable when filling the database and to precisely follow the selected data type, since further processing of variables in many currently used statistical programs may depend on it.

In the practice of processing the results of the research two types of statistical data analysis are used - primary (planned) and secondary (unplanned).

Primary data analysis - is used to study and describe the laws, the existence of which is assumed by the researcher, and which are the actual hypothesis of the study. In this case the characteristics are analyzed, the study of which is taken into account when planning the study, and pre-formulated hypotheses are tested.

Secondary data analysis - used to form prospects of the study, search, exploration of potential laws and hypotheses. In such a case, "screening" of unscheduled data in a particular job is performed, which is often advisable at the first stage of familiarization with the data.

One of the main components of any data analysis is descriptive statistics. Its main task is to provide a concise and concentrated characteristics of the phenomenon under study in numerical and graphical form.

The population value of the parameter (mean value, median, share, etc.) cannot be obtained (an exception is when the study is conducted on a group that includes all members of the population). However, the population value of the parameter can be estimated by

sampling. The accuracy of this estimate depends on the measurement method (measurement error), the size and representativeness of the sample (sample error), and biological variation.

Indicators of descriptive statistics can be divided into several groups:

- position indicators describing the position of the experimental data on the numerical axis. Examples of such data are the maximum and minimum elements of the sample, mean value, median, mode, etc.:

- scatter indicators describing the degree of data scatter;

relative to the central trend. These include: sample variance, the difference between the minimum and maximum elements (range, sampling interval), and others:

- asymmetry indicators: the position of the median relative to the average, etc.:

- graphical representations of the results - histogram, frequency chart, etc.

When using descriptive statistics, it is important to consider the type of data and distribution parameters, which are characterized by asymmetry indicators and a distribution histogram. The most commonly used criteria for testing the hypothesis about the distribution law are the Pearson criterion, the χ^2 criterion and the Kolmogorov-Smirnov criterion: if the trait distribution differs from the normal one

distributions with statistical significance less than 0.05 ($p < 0.05$) distribution of a trait in the sample is considered abnormal, and vice versa.

The main types of distribution of signs are: discrete (for discrete signs - binomial, Poisson distribution, Bernoulli distribution) and continuous (for continuous signs - normal (Gaussian, or Gaussian distribution), lognormal, exponential, chi-square χ^2). According to the type distribution applies two principles of statistical processing: parametric and non-parametric. The parametric principle includes all methods for analyzing normally distributed quantitative traits. The non-parametric principle is used in all other cases - for analyzing quantitative traits regardless of the type of their distribution and for analyzing qualitative traits.

Non-parametric methods are considered less powerful than parametric ones, i.e. sometimes they do not reveal statistical patterns that can be detected using parametric methods. At the same time, non-parametric methods are more reliable in cases where there is a doubt that the analyzed sign has a normal distribution. For normally distributed features, parametric and non-parametric methods give similar results.

The indication in the data view measures the central trend (average, median, mode) automatically informs the reader about the normal distribution of the trait. In a normal distribution, all three indicators more or less coincide, and in an asymmetric distribution, they are not.

Mode (Mo) is the most frequent value in the sample, or the average value of the class with the highest frequency. Fashion as a central trend is most often used to give a general idea of distribution. In some cases of distribution, there may be two modes, in which case it indicates a bimodal distribution, which indicates the presence of two relatively independent groups.

The median (Me, Md) corresponds to the central value in a consistent series of all obtained values or the average value of the most frequently encountered sample values. The median along with quartiles is used to represent discrete variables or quantitative continuous variables with abnormal distribution.

The arithmetic average (M) is the index of the central trend, obtained by dividing the sum of all data values by the number of these data. The arithmetic average is used to represent quantitative variables with a normal distribution. The average value, as a measure of the central tendency in descriptive statistics of quantitative data, has one of two representations. The first is in the form of "M ± S", or in the foreign tradition M (S), where M is the mean, and S is the standard deviation. The standard deviation is intended to describe samples with a normal distribution and is not adapted for distributions other than normal. With a normal distribution, the range M ± S fits about 70% of all values of the trait.

The second presentation of the results is in the form "M ± m", where m is Standard Error of Mean, defined as follows: $m = s / \sqrt{n}$.

However, this form of presentation of data in medicine is uninformative. The use of standard error of the mean is used in physics, where, when measured parameters of identical objects, the variability of the results is determined only by random errors, and with an increase in the number of measurements, one can get a mean value closer to the true one, with a smaller standard error of the mean. In medicine, the objects of observation are complex systems that differ significantly in their properties, which determines the practical absence of the true value of the parameter. In fact, in biology (and, respectively, in medicine) it is not the exact value that is determined, but the range in which most of the values of the trait under study fit: width of distribution. Therefore, the optimal description of the width of the distribution in medical research is currently taken to represent the 95% confidence interval, indicating the lower (5%) and upper (95%) boundaries.

The confidence interval is a range of values, which with a certain researcher's probability (most often in medicine, this is $\alpha = 0.05$ or 95%) includes the real population value.

The most adequate non-parametric characteristic of the width is quantiles. Quantiles represent the frequency of hitting variable values at certain intervals. The most commonly used division is 10 (10% each) or by 4 intervals (25%, 50%, 75%). When divided into four quantiles (referred to as quartiles), three numbers are enough to provide an estimate of the central trend, width and asymmetry of the distribution of results: the bottom quartile (25%), 50% quartile, which corresponds to the median, and the top quartile (75%). This method of providing data is one of the most compact and convenient.

The task of inductive statistics is to test statistical hypotheses about the distribution law, and the main area of application is the use in biomedical research to compare two different samples for belonging to the general population. The belonging of two samples to one general aggregate indicates the absence of differences between them.

For this, statistical hypotheses are formulated:

- H_0 hypothesis about the absence of differences (null hypothesis);

- H_1 hypothesis about the significance of differences (alternative hypothesis).

That is, it is necessary to solve the question of the randomness of the identified differences, the decision on whether the identified differences are evidence of a different state and / or evidence of the effect of the intervention depends on it. The quantitative characteristic of chance is represented by probability theory in the form of a *p-value*. Than this value the greater, the greater the likelihood of no difference in favor of the null hypothesis, and the smaller it is, the greater the likelihood of differences in favor of an alternative hypothesis.

p-value is a quantitative characteristic only statistical, but not of clinical significance. In the presence of statistical significance, it is necessary to decide on the clinical importance of the identified differences. This is especially true of secondary data analysis, unplanned. In the initial planned analysis, the data is usually checked for the statistical significance of clinically important differences.

Probability theory basically uses the concept of an admissible error, and the error is a necessary component of statistical analysis that affects the p-value. The allowable error level on which the p-value depends is chosen by the researcher. In biomedical in research, it is customary to use two types of errors: an error of the first kind, which corresponds to the notion of the level of statistical significance α (alpha), and an error of the second kind β (beta), which corresponds to the notion of statistical power 1-β.

Error of the first kind (significance level α) – validity mistaken recognition of differences, that is, an alternative hypothesis. In biomedical research, the level of 0.05 is traditionally chosen as a critical threshold of significance, which allows for the presence of an error of the first kind 5 times in 100 comparisons. When $p \leq \alpha$, differences are accepted statistically significant. And the smaller the p-value, the less such errors: for example, when $p = 0.01$, it is considered that the error of the first kind is possible 1 time in 100 comparisons, at $p - 0.001$ - one time in 1000 comparisons. However, in exploratory pilot studies, a significance level of $\alpha = 0.1$ is allowed, to identify emerging differences and / or relationships for the purpose of further planning on the basis of their new research with sufficient significance.

Second type error β (statistical power 1-β) - the admissibility of the erroneous rejection of the presence of differences or, which is the same, the erroneous recognition of the absence of differences, respectively, the erroneous recognition of the null hypothesis, due to the lack of data The second kind error is expressed in the form of statistical power equal to 1- β. Power is needed to determine the sufficiency of the sample size, especially when proving the absence of statistical significant differences in bioequivalent studies. With adequate statistical power lack of statistical significant differences is indeed recognized as such. With inadequate power it is impossible to assert about the equivalence (similarity) of groups. In biomedical research, the error value is taken as the critical threshold the second kind of $\beta = 0.1$ or $\beta = 0.2$, which in the form of statistical power, expressed as a percentage, is 90% or 80%, most often - 80%: the probability that out of 100 in 80 cases a really existing difference will be identified and in 20 cases - missed.

The choice of criteria depends on the task, the type of data and number of measurements. Thus, for quantitative data with distributions close to normal, parametric methods are used, based on such indicators as mean value and standard deviation. For comparison unpaired *t-criterion* is used for two independent samples, paired *t-criterion* is used for two dependent samples.

When processing small samples (less than 16 objects, at which the distribution begins to differ significantly from normal), non-parametric methods are used to compare non-quantitative data — the *Mann-Whitney U-test* for two independent samples, the Wilcoxon test for comparing two dependent samples, the χ^2 test (hi-square) to test the statistical hypothesis about the presence of a connection between two qualitative signs.

Fisher's exact test

The χ^2 criterion is suitable for analyzing contingency tables of 2x2, if the expected values in any of its cells are not less than 5. When the number of observations is small, this condition is not satisfied and the criterion χ^2 not applicable. In this case, Fisher's exact test is used. It is based on the enumeration of all possible options for filling the contingency table for a given number of groups, therefore, the smaller it is, the easier it is to apply it.

Correlation - the relationship between two or more variables (in the latter case, the correlation is called multiple or cumulative). The purpose of the correlation analysis is to establish the presence or absence of this interconnection In the case when there are two variables, the values of which are measured in the ratio scale, the Pearson linear correlation coefficient r is used, which takes values from -1 to +1 (its zero value indicates the absence of correlation).

The term "linear" indicates that the presence of a linear relationship between variables is investigated.

For data measured in the ordinal scale. the Spearman's rank correlation coefficient should be used, since it is non-parametric and detects a trend — changes in variables in one direction, denoted by r_s, and determined by comparing ranks — numbers of values of compared variables in their ordering. Spearman's correlation coefficient is less sensitive than Pearson's correlation coefficient. It is important to note that the value of the correlation coefficient close to plus one or minus one indicates the strength of the relationship of variables, direct or inverse. but does not say anything about the cause-effect relationship between them.

In contrast to the correlation analysis, regression analysis - not only indicates the existence of a relationship between an independent variable and one or several dependent variables. but also allows you to determine this dependence quantitatively. Independent variables are called regressors or predictors, and dependent variables are called criterion. Again. the terminology of dependent and independent variables reflects only the mathematical dependence of variables. and not cause-effect relationships. There are several types of linear and nonlinear regression analysis, which allow detecting mathematical dependence between several variables. however all these methods are parametric. which makes it impossible to use them for processing qualitative data. The non-parametric analogue of multiple regression is logistic regression with two gradations of the dependent trait (binary logistic regression) and more (multinominal logistic regression).

4. Data and Methods

In this thesis work, data of children mortality rate of Kazakhstan in year 2005-2018 was used. It was obtained from the annual reports and statistical year book of UNESCAP (2000-2018), UNICEF (2000-2018) and WHO (2005-2018) and the Ministry of National Economy of the Republic of Kazakhstan used the Statistics Committee, which reflects the mortality rate and the values of its coefficients for the considered causes of death of children in Kazakhstan for 2005-2018. (www.stat.gov.kz).

Mortality of children in age groups, perinatal death, infant death, mortality of children under five years old and higher were taken. Mortality of children was also observed by region in 2009–2016 years.

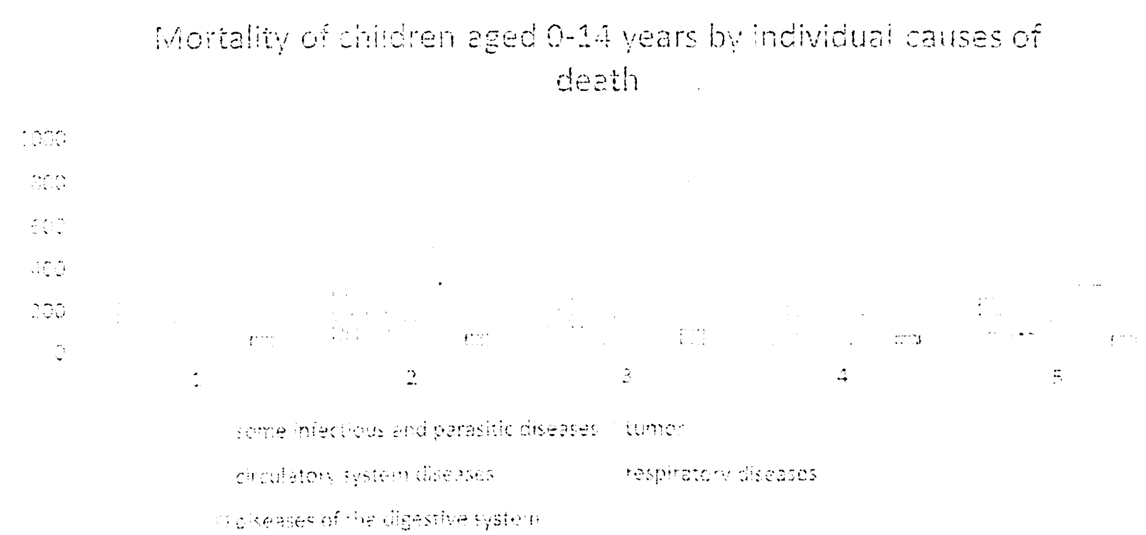

Diagram 1. Mortality of children aged 0-14 years by individual causes of death in Kazakhstan 2009-2016 years.

The above chart shows the mortality of children aged 0-14 years for individual reasons in 2009 and 2016. As can be seen from the diagram in Kazakhstan, most of all children die from respiratory diseases, infections and from parasitic and at the site we have a tumor, but most of all children die from respiratory diseases.

These diseases are associated with exposure to the respiratory tract infections in the form of viruses, bacteria and even fungi, which causes the characteristic symptoms of a cold. Children most often suffer from acute respiratory infections, especially in the first 3 years of their lives.

After birth, the baby faces a huge number of microbes in its environment, but he has not yet learned how to defend against them. The body of the fetus receives the necessary antibodies through the placenta of the mother, and in the prenatal period it does not occur with bacteria. After birth, all the necessary nutrients, including immunoglobulins that protect against infection, the child receives through breast milk. However, this is not enough in the first years of life, when the child is actively exploring the surrounding, and therefore is faced with a large number of microbes. Therefore, in the first 3 years of life, ARD is the most common disease in a child.

The next outbreak of the disease occurs at 4-5 years. when the immunoglobulins obtained from the mother's body begin to disintegrate. The first immunoglobulins that the child's body begins to produce on its own are not strong enough to protect it from various types of infections.

According to the statistics committee of Kazakhstan, most children die in infancy. under five years and perinatal period.

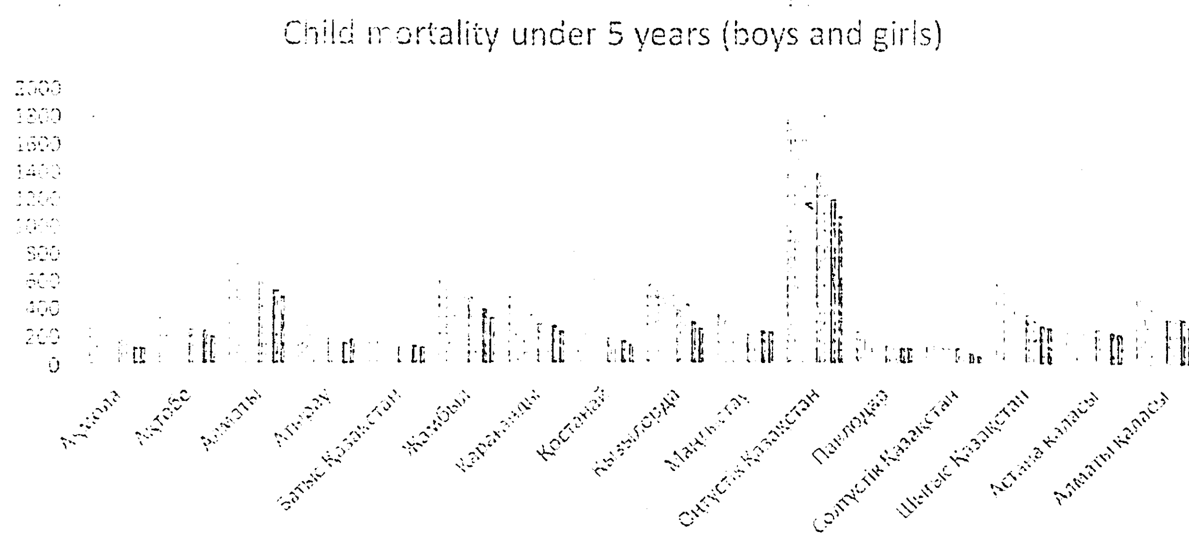

Diagram 2. Child mortality under 5 years by region in Kazakhstan 2009-2016 years.

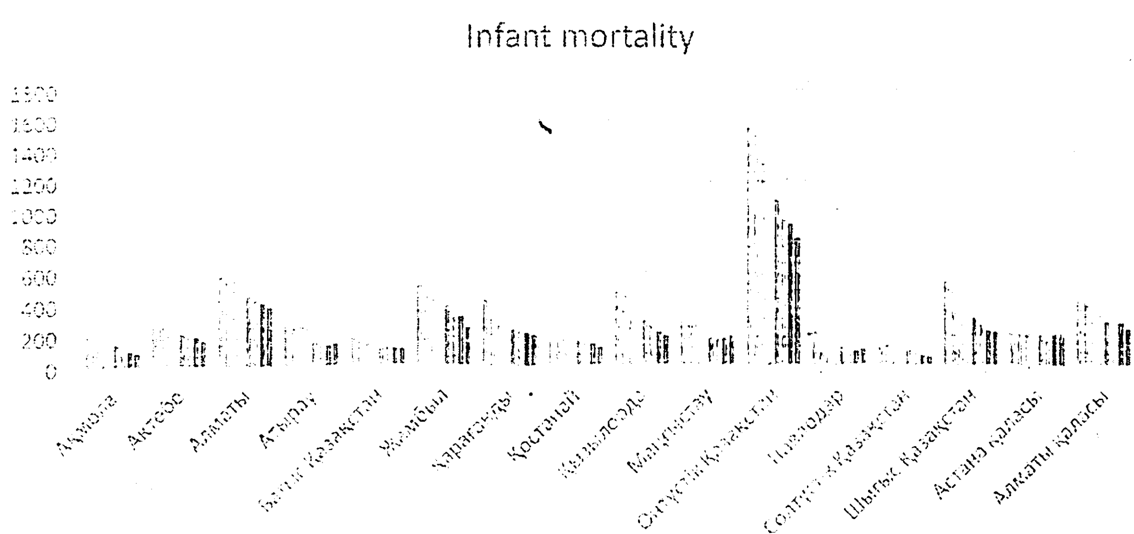

Diagram 3. Infant mortality by region in Kazakhstan 2009-2016 years.

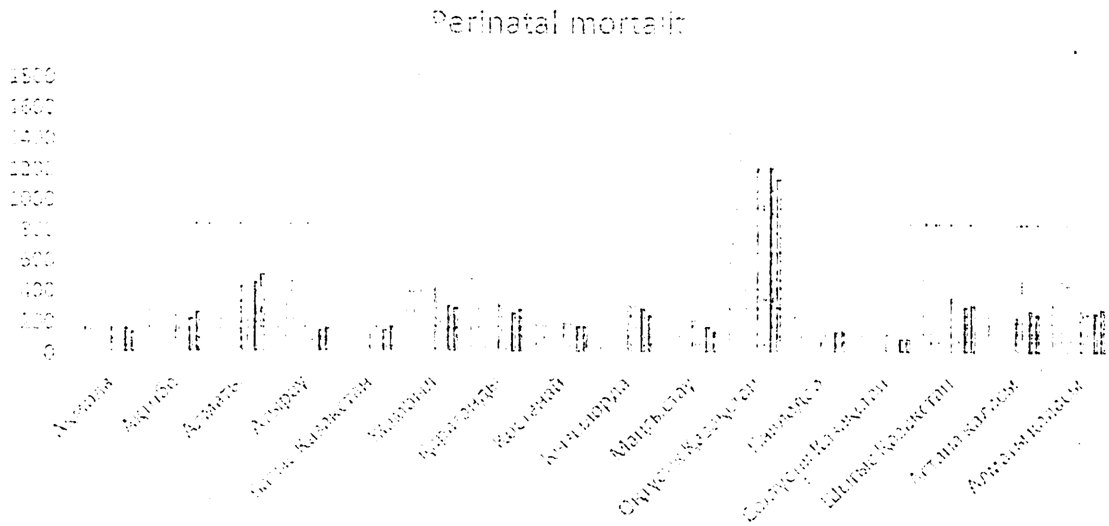

Diagram 4. Perinatal mortality by region in Kazakhstan 2009-2016 years.

To assess the state of public health, quality of life of the population, degree of development and organization health services in different countries, WHO recommendations along with traditional criteria (total, maternal, perinatal and infant mortality) use the child mortality rate in age those up to 5 years. One of the main indicators of the state public health, including childhood, is death rate [4,7].

Considering the child mortality rates, it should be noted that most children die in infancy. Infant mortality is one of the demographic factors that most vividly reflect the level of a country's development and the economic and social changes taking place in it.

Currently, the mortality rate of children aged 0 to 5 years in the Republic of Kazakhstan is higher than that in comparison with CIS countries that have similar socio-economic and medical-organizational conditions, which is of interest to Kazakhstani demographers.

The infant mortality rate in Kazakhstan is higher than in Belarus, Russia, Ukraine [5]. As noted during the meeting on the introduction of effective technologies in Kazakhstan under the WHO programs, infant mortality in Kazakhstan is very high - 20.4 per 1000 newborns and for this indicator it takes the 2nd place after Kyrgyzstan [6].

According to the Statistics Agency of the Republic of Kazakhstan for 2012 in the republic annually die over 5 thousand children under the age of 1 year (14.84 per 1000 live births).

Now we will consider infant mortality in Kazakhstan for individual reasons.

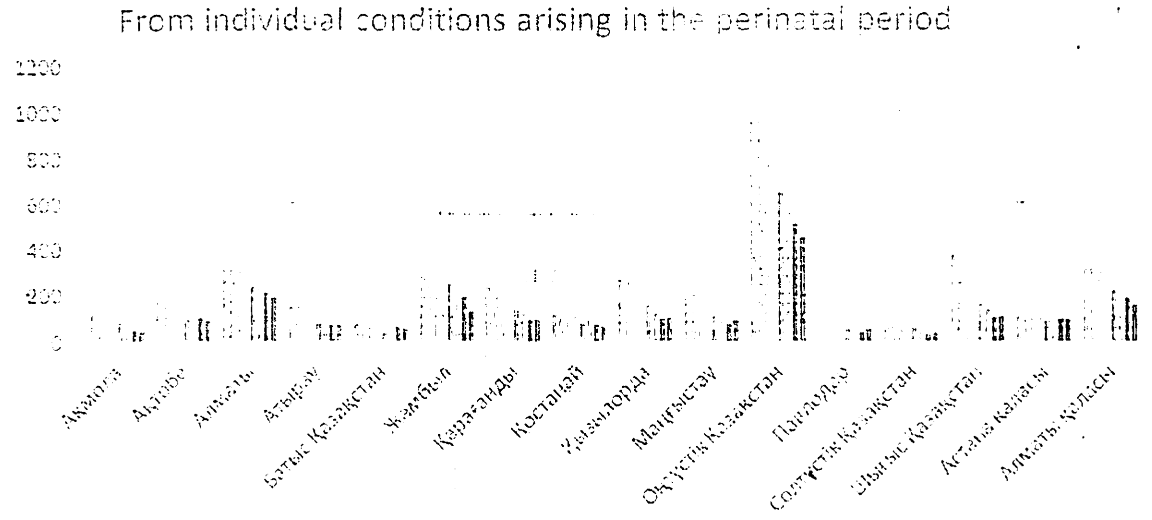

Diagram 5. From individual conditions arising in the perinatal period by region in Kazakhstan 2009-2016 years.

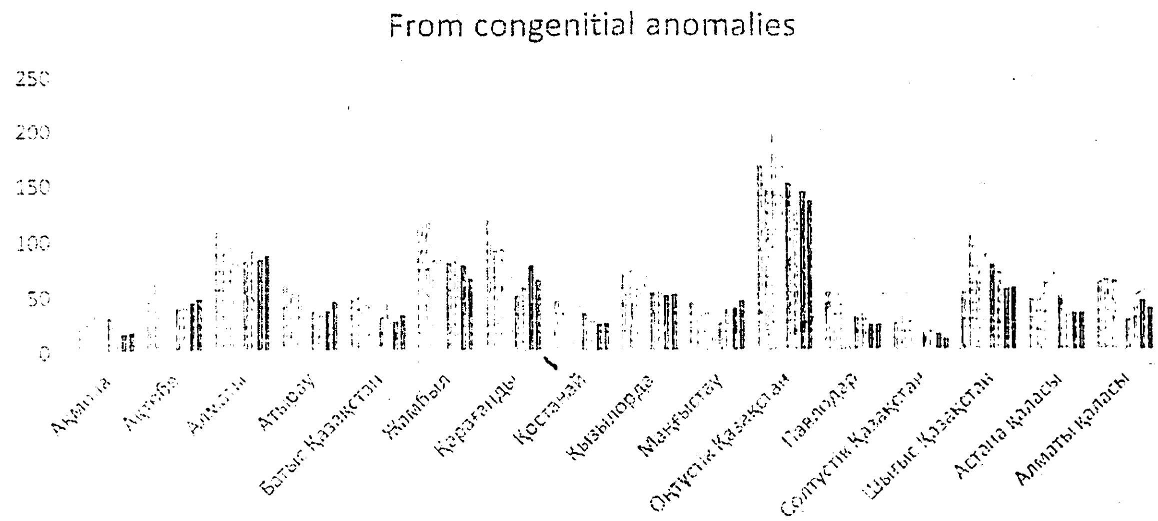

Diagram 6. From congenital anomalies by region in Kazakhstan 2009-2016 years.

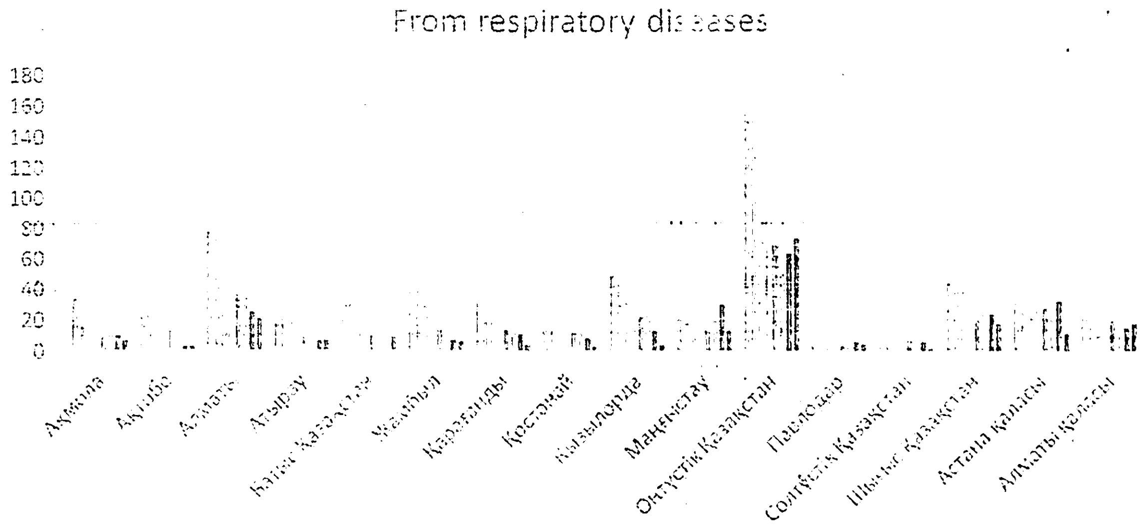

Diagram 7. From respiratory diseases by region in Kazakhstan 2009-2016 years

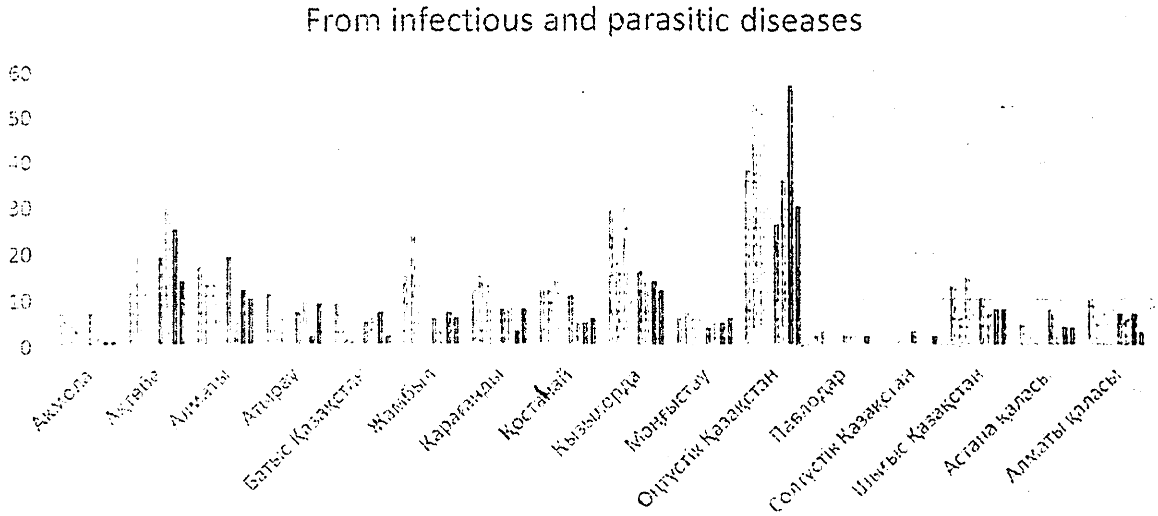

Diagram 8. From infectious and parasitic diseases by region in Kazakhstan 2009-2016 years

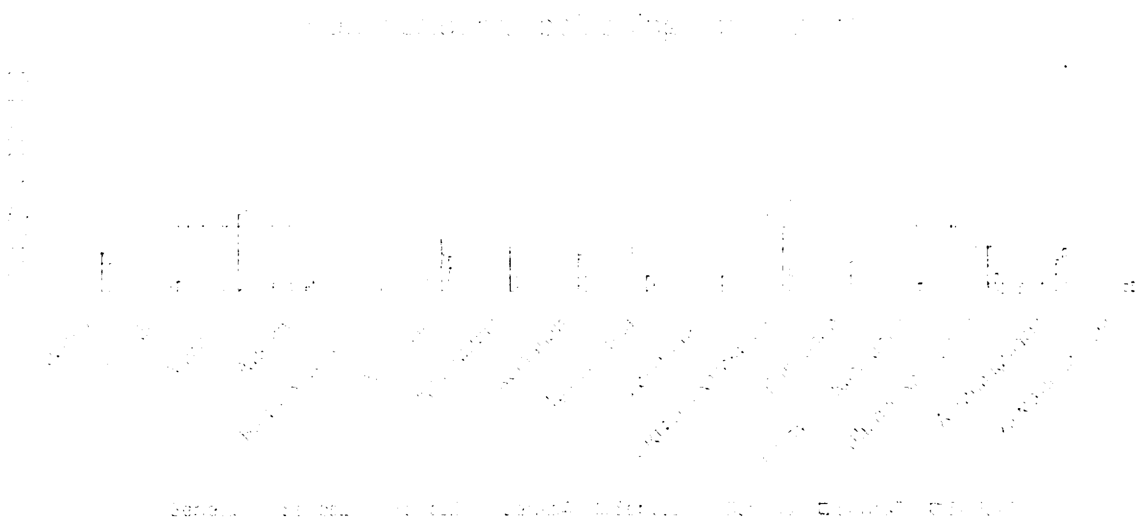

Diagram 9. From accidents. poisoning and injuries by region in Kazakhstan 2009-2016 years.

The data of independent studies (the program of medical and demographic research of Kazakhstan) conducted in Kazakhstan in 1995 and 1999 and the multi-indicator cluster survey conducted in 2006 showed that over 15 years (from 1989 to 2004) the infant mortality rate increased from the initial level (1989). amounting to 55.0% to 62% in 1999. and then sharply falling to 34.0 stabilized at a relatively high level.

The decline in infant mortality occurred mainly due to a decrease in fertility. as they are demographic indicators. and there is a direct correlation between them ($r = 0.7$, $p < 0.01$).

Since 2008 with the introduction of live birth criteria and stillbirth the infant mortality rate was 20.7. and in 2009 there is a tendency towardsdecrease to 18.4 per 1,000 live births.

Perhaps the reason for such a significant difference is the use of different criteria for live birth and stillbirth. At the same time. the Ministry of Health used the criteria of the Soviet period. the program of medical and demographic research of Kazakhstan - international. recommended by WHO [5.6.10].

Different approaches to the definition of live birth leads to a distortion of the real infant mortality rate in the Republic of Kazakhstan by registering some of the newborns as late miscarriages or stillborn. It is especially necessary to pay attention to one trend of infant mortality in our country - to the increasing role of the Southern region. where infant mortality is consistently high. If in 1990 among all the children who died in the Republic. in the first year of life. children in the South Kazakhstan region accounted for 32.9%. in 2000 it was 36.7%. At the same time. out of all those born in Kazakhstan. the proportion of those born in this area grew significantly less from 24.2% to 26.8% [8].

Despite a significant reduction in specific the weights of exogenous causes of infant mortality. they retained crucial in the structure of the indicator [9]. In the Southern region. features of the season fluctuations in infant mortality in the first year of life.

If in the republic the infant mortality is somewhat higher in July, August, January and February. then in South Kazakhstan region peak indicator accounted for June. July. August and September. Therefore. this circumstance the property requires attention to the summer peak infant mortality.

Success in reducing it will be depend on how effective will be measures against seasonal outbreaks of intestinal infections in region.

Along with this. in the 21st century. the pollution of the external environment and the environmental crisis that are observed throughout the world are currently tending to increase. Ecologically unfavorable environment has a direct impact on the health of children. The growth of environmentally-related diseases such as asthma. acute respiratory infections. diarrhea, perinatal diseases, and congenital malformations has been noted [9.10].

The study of mortality and morbidity of children from 0 to 5 years in environmentally disadvantaged conditions cities and districts of Semipalatinsk test landfill as well as environmentally unfriendly Priaralye. dynamics of demographic indicators. in whole. reflected the unfavorable situation in relation to the health of the child population due to unstable infant mortality rates (15.2 ‰ -23.6 ‰). an increase in the overall mortality of the population (13 ‰ - 14.6 ‰). In the country. the last index has relatively stable level (10.0‰-10.4‰).

An analysis of the data for 2007–2009 showed that the mortality of children under 1 year of respiratory diseases and pneumonia ranks 3rd among all causes of death, and from infectious diseases 1 place. In 2008. in the Republic of Kazakhstan. 33,774 cases of pneumonia were registered in children under 5 years of age. The total number of deaths of children under 5 years old in 2008 was 8.225 children; about 1.500 children died from pneumonia.

Now I will give an example of the above linear regression. For this, I covered the population of Kazakhstan. where the mortality of children is most pronounced. This data is processed using the MK Exzel program. In this study, data of child mortality rate of Kazakhstan in year 2005-2018 was used. It was obtained from the annual reports and statistical year book of UNESCAP (2000-2018). UNICEF (2000-2018) and WHO (2005-2018) and the Ministry of National Economy of the Republic of Kazakhstan used the Statistics Committee. which reflects the mortality rate and the values of its coefficients for the considered causes of death of children in Kazakhstan for 2005-2018. (www.stat.gov.kz).

First. I will consider the southern part of the Kazakhstan region. According to the Statistics Committee of Kazakhstan, children die here most of all. I took from 2009 to 2016 the data on child mortality. This I showed the above diagrams. How y - saw the overall infant mortality rate of children. X_1 - From individual conditions arising in the perinatal period, X_2 -From congenital anomalies, X_3 - From respiratory diseases, X_4 - From infectious and parasitic diseases, X_5 - From accidents, poisoning and injuries.

Table 1. Infant mortality in South Kazakhstan

	Y	X1	X2	X3	X4	X5
2009	1547	1030	167	166	38	68
2010	1521	1008	145	158	53	60
2011	1329	829	196	87	51	51
2012	1211	731	167	93	42	44
2013	1071	650	151	69	26	52
2014	938	563	129	50	36	38
2015	913	508	143	67	57	25
2016	813	450	135	73	30	35

Table 2. Correlation between variables

	Y	X1	X2	X3	X4	X5
Y	1					
X1	0.807208	1				
X2	0.779311	0.731375	1			
X3	0.892953	0.803251	0,693072	1		
X4	0.727063	0.792076	0.786982	0.756525	1	
X5	0.894374	0.782232	0,754454	0.814356	-0.1078	1

Here we see a correlation between variables. I want to test them for multicolliniarity. If the result of multicolliniarity is greater than or equal to 10, then there are problems in our variables. I will do this only on the results of correlations. Multicolliniarity is checked by the formula:

$$VIF = \frac{1}{tolerance}$$

$$Tolerance = 1\text{-}R^2 \qquad [19]$$

We know that if a VIF model is greater than 10, then the multicolliriarity between coefficients is not. This brings us to the fact that the model will not be correct. In Table 2 we see the correlation between the variables. If we calculate them, then all variables less than 10 show that the multicoliriarity of intern variables is there and we can continue building our model.

Now we can consider the regression analysis of the Table 1.

Table 3. Variance analysis for infant mortality in South Kazakhstan.

R^2	F
0.87	0.00638

The accuracy of the significance level of the Fisher test (significance F) is significantly less than 0.05, which means the model is significant.

The degree of accuracy of the process model description of the R-square is 0.87, which indicates a high approximation accuracy (the model describes the process well).

Table 4. Variance analysis for infant mortality in South Kazakhstan.

	Coefficient	p-value
Y	81,8215	0.16237
X_1	0.6944	0.03608
X_2	0,36635	0.01355
X_3	0,16261	0.03173
X_4	-0,7578	0.04431
X_5	-2.8824	0.03222

The p-value for the coefficient X_1 is less than 0.05, which means this coefficient can be considered non-zero.

The p-value for the X_2 coefficient is less than 0.05. which means this coefficient can be considered non-zero.

The p-value for the X3 coefficient is less than 0.05. which means this coefficient can be considered non-zero.

The p-value for the X_4 coefficient is less than 0.05. which means this coefficient can be considered non-zero.

The p-value for the X_5 coefficient is less than 0.05. which means this coefficient can be considered non-zero.

The value of the free term (Y-intersection) 81.8215.

From here you can write a model of infant mortality in South Kazakhstan

$$Y = 0{,}6944X_1 + 0{,}36635X_2 + 0{,}16261X_3 - 0{,}7578X_4 - 2{,}8824X_5 + 81{,}8215 + \varepsilon$$

Here X_1 - From individual conditions arising in the perinatal period, X_2 - From congenital anomalies, X_3 - From respiratory diseases, X_4 - From infectious and parasitic diseases, X_5 - From accidents, poisoning and injuries, Y – value of the free term, ε- the remainder.

We see that infant mortality in South Kazakhstan is 69% derived from individual states occurring in the pre-natal period. What is individual states occurring in the pre-natal period:

- Damage to the fetus and newborn caused by the mother's condition, complications of pregnancy, childbirth and delivery
- Disorders related to the duration of pregnancy and fetal growth
- Birth injury
- Respiratory and cardiovascular disorders characteristic of the perinatal period
- Infectious diseases specific to the perinatal period
- Hemorrhagic and hematological disorders in the fetus and newborn
- Transient endocrine and metabolic disorders specific to the fetus and newborn
- Disorders of the digestive system in the fetus and newborn
- Conditions affecting the skin and thermoregulation in the fetus and newborn
- Other disorders arising in the perinatal period.

Now consider the quality of medical service for infant mortality. For this, I took the data on the birth rate of children. the number of beds, the number of doctors, the average health care personnel.

Y-infant child mortality

X_1 –birthrate

X_2 – number of beds

X_3 – number of doctors

X_4 – average health care personnel

Table 5. Infant mortality

	Y	X1	X2	X3	X4
2009	1547	77865	2910	7166	19064
2010	1521	77990	2914	8039	20561
2011	1329	78456	3001	8492	23495
2012	1211	78516	2691	8231	25845
2013	1071	78998	2591	8295	26385
2014	938	79115	2828	9280	24208
2015	913	81167	2866	8795	24077
2016	813	80860	2860	9315	25433

Table 6. Correlation between variables

	Y	*X1*	*X2*	*X3*	*X4*
Y	1				
X1	-0,87249	1			
X2	0,830616	-0,86079	1		
X3	-0,8758	0,719124	0,74005	1	
X4	-0,77	0,617922	-0,6495	0,662957	1

We know that if a VIF model is greater than 10. then the multicolliriarity between coefficients is not. This brings us to the fact that the model will not be correct. In Table 6 we see the correlation between the variables. If we calculate them, then all variables less than 10 show that the multicoliriarity of intern variables is there and we can continue building our model.

Now we can consider the regression analysis of the Table 5.

Table 7. Variance analysis for infant mortality

R^2	F
0,86	0,01438

The accuracy of the significance level of the Fisher test (significance F) is significantly less than 0.05. which means the model is significant.

The degree of accuracy of the process model description of the R-square is 0.86. which indicates a high approximation accuracy (the model describes the process well).

Table 8. Variance analysis for infant mortality

	Coefficient	p-value
Y	9712,504	0,033258
X_1	-0,10902	0,048196
X_2	0,638003	0,020932

X_3	-0.21368	0.047365
X_4	0,3361	0,038861

The p-value for the coefficient X_1 is less than 0.05, which means this coefficient can be considered non-zero.

The p-value for the X_2 coefficient is less than 0.05, which means this coefficient can be considered non-zero.

The p-value for the X3 coefficient is less than 0.05, which means this coefficient can be considered non-zero.

The p-value for the X_4 coefficient is less than 0.05, which means this coefficient can be considered non-zero.

From here you can write a model of infant mortality

$$Y = 9712{,}504 - 0{,}10902x_1 + 0{,}638003x_2 - 0{,}21368x_3 + 0{,}3361x_4$$

Most of the percent is the second coefficient – 63%. This means that in South Kazakhstan there is a lack of places in medical institutions. In our mathematical model, the second coefficient X_2 means the number of beds. In second place is the average number of medical personnel – 33%.

Conclusion

In this work constructing of a model equation for the dynamics of children mortality in population of Kazakhstan is described. Mortality is an important aspect of demographic process, closely linked to the socio-economic prospects for the development of society. The mortality rate is also connected with the gender aspect, age structure of countries population, and also in lifespan evaluation. Highlighted and described factors that have the greatest impact on the mortality rate of children of the population of Kazakhstan. The regional peculiarities of this process are considered. In particular, the study has an impact on the lack of qualifications and poor quality of medical personnel and non-compliance with the level of medical service. In addition, another important factor is the dissatisfied state of ecology with man-made causes. the causes of mortality in the context of regions are investigated and complex of factors that influenced the differentiation of mortality among the regions of Kazakhstan are analyzed as well as the dynamics of infant mortality.

$$Y = 0{,}6944X_1 + 0{,}36635X_2 + 0{,}16261X_3 - 0{,}7578X_4 - 2{,}8824X_5 + 81{,}8215 + \varepsilon$$

We see that infant mortality in South Kazakhstan is 69% derived from individual states occurring in the pre-natal period.

$$Y = 9712{,}504 - 0{,}10902x_1 + 0{,}638003x_2 - 0{,}21368x_3 + 0{,}3361x_4$$

Most of the percent is the second coefficient – 63%. This means that in South Kazakhstan there is a lack of places in medical institutions.

References

[1]. Nazarbayev N.A. Kazakhstan - 2030: prosperity, security and improvement of the welfare of all Kazakhstanis. Message of the President of the country to the people of Kazakhstan. - Almaty: Bilim. 96 p.

[2]. Nazarbayev N.A. The future well-being of the country is to start building today. - Almaty: Kazakhstanskaya Pravda. 1997. October 11. C.2 –

[3]. Shokamanov Yu.K. Trends in human development in Kazakhstan. - Almaty. 2001. P.143.

[4]. Demographic Yearbook of the USSR 1990 Goskomstat of the USSR. - M .: Finance and Statistics, 1990. P.367

[5]. Demographic Statistical Yearbook of Kazakhstan. 1999. Art. Sat - Almaty: Agency of the Republic of Kazakhstan on Statistics. 2000. - C5.41.24.

[6]. Statistical Yearbook of Kazakhstan. Art. Sat - Almaty: National Statistical Agency of the Ministry of Economy and Trade of the Republic of Kazakhstan, 1997. - P.21.19.20.

[7]. Statistical Yearbook of Kazakhstan in 1990 - Alma-Ata: Kazinformcenter. –C.43.

[8]. Concise Statistical Yearbook of Kazakhstan 1995: Statistical collection of the State Statistics Committee of the Republic of Kazakhstan. - Almaty: Kazinformtsentr, 1996. –C.19.

[9]. Summary statistical yearbook of Kazakhstan, 1998: Statistical compilation Agency of the Republic of Kazakhstan on Statistics. - Almaty: Kazinformcenter, 1999. P.10.

[10]. Statistical Yearbook of Kazakhstan. Art. Sat Almaty: Agency of Statistics of the Republic of Kazakhstan, p.22.

[11]. K. Kozhumov. Health: expenses increase, but "sores" still remain. AIF Kazakhstan, № 48, 2010.

[12]. Reproduction of the USSR population / A.Г. Volkov. V.A. Belova, G.A. Bondarskaya and others. M .: Finance and Statistics, 1983. p.1Ť3.

[13]. Rakhmatullin E.V. Medical and social aspects of fertility and reproductive behavior of women in municipal area: dis. ... Cand. honey. Sciences: 14.00.33. - M., 2008. - 152 p.

[14]. Demographic Yearbook of the Regions of Kazakhstan. Statistical collection. - Almaty, 2006, 2007. 2009. 2012 - 548 p., 392 p., 637 p.

[15]. Vaganov N.N. Health protection of women and children in modern conditions // Medico organizational problems protection of motherhood and childhood. Digest of articles. - 2001. - pp. 129-132, 136-142.

[16]. Baranov A.A. Children's infections as a cause of infant and child mortality // Pediatrics. - 1991.-№6. C.5-9.

[17]. Bogdanovich N.S., Vanilovich I.A. Maternal and Child Health: Inform. Analyte. materials. - M.: UNIPAK, 2002. - 16 p.

[18]. Isaev DS, Nadirova K.I. Reserves to reduce infant mortality and ways to implement them // Pediatrics.- 1998.- №1.- P.38-40.

[19]. Modeling Asia's Child Mortality Rate: A Thinking of Human Development in Asia. 7th International Economics & Business Management Conference, 5th & 6th October 2015. Available online at www.sciencedirect.com

[20]. https://www.who.int/gho/child_health/mortality/ChildCME_method.pdf

[21]. https://www.unicef.org/kazakhstan/

[22]. ОСТРЫЕ РЕСПИРАТОРНЫЕ ЗАБОЛЕВАНИЯ У ДЕТЕЙ ШКОЛЬНОГО ВОЗРАСТА: ОСОБЕННОСТИ ТЕЧЕНИЯ И РЕЦИДИВИРОВАНИЯ
В ЗАВИСИМОСТИ ОТ КОНСТИТУЦИОНАЛЬНОГО ТИПА УДК 616.9-053.2-084:578.831.31

[23]. Main tendencies of infant mortality in Kazakhstan УДК: 616-036.88-053.2-02(574)