Linear Regression Approximate Models for Predicting Severe Course of Bronchial Asthma

Oleh Pihnastyi [1], Olha Kozhyna [2] and Tetiana Kulik [2]

[1] National Technical University "Kharkiv Polytechnic Institute", 2 Kyrpychova, Kharkiv, 61002, Ukraine
[2] Kharkiv National Medical University, 4 Nauky Avenue, Kharkiv, 61022, Ukraine

Abstract

This paper discusses using of a multifactor linear regression model to predict bronchial asthma severity. The study is aimed to develop the method of some five-factor linear regression approximate models building and to substantiate the areas of their use. 142 factors obtained during workup of 90 children at the age from 6 to 18 were analyzed. 70 children with bronchial asthma of various degrees of severity as well as 20 healthy school-aged children were included into the main group. The degree of qualitative and quantitative factors association being studied was determined to select the predictors having an effect on severe course of bronchial asthma. The following factors were used to build an approximate linear regression model: bronchial asthma in relatives of second generation, atopic dermatitis, allergic rhinitis, sheep wool, rabbit hair, domestic dust, severe. A comprehensive study of the value under investigation as well as its dependence on a big number of factors is the basis of the proposed method. The developed technique of linear regression multi-parameter models building allows us to simplify the process of linear regression models building which can be used both for preliminary asthma severity prediction and for detailed study of its course.

Keywords

Bronchial asthma, child, regression model, severe asthma, prediction, allergic rhinitis, atopic dermatitis

1. Introduction

Allergic diseases prevalence in both children and adults increases year by year. Bronchial asthma holds a specific place among allergic diseases. Currently, due to the rapid increase of asthma incidence, this pathology can be considered as a global medical problem. According to GINA, in 2019, asthma attacked 262 million people and caused death of 461000 people [1]. Bronchial asthma in children is the most common chronical respiratory pathology. It is known that an early disease manifestation and intensity of clinical symptoms are defined by the combination of genetic and environmental factors. Multiple asthma manifestations were studied and their connection with various pathogenic mechanisms was revealed [2]. Depending on the disease severity at an early age, the risk of a bronchopulmonary pathology onset in adult life increases. Bronchial asthma wields major influence on a life quality of patients, it leads to significant economic losses [3, 4]. A comprehensive study of the disease factors will allow not merely to define a risk group among children but also to influence the disease severity. Research in genetics, etiology, pathogenesis of asthma has expanded knowledge in this area, but there is still a significant number of patients with severe asthma characterized by poor control [5]. The diagnosis of asthma in a midchildhood is complicated due to the similarity of clinical manifestations with other diseases. In applied medicine, predictive models based on anamnestic data are used to

ITTAP'2022: 2nd International Workshop on Information Technologies: Theoretical and Applied Problems, November 22–24, 2022, Ternopil, Ukraine
EMAIL: pihnastyi@gmail.com (A. 1); olga.kozhyna.s@gmail.com (A. 2); tv.kulik@knmu.edu.ua (A. 3)
ORCID: 0000-0002-5424-9843 (A. 1); 0000-0002-4549-6105 (A. 2); 0000-0002-8842-892X (A. 3)

improve the diagnosis of asthma at an early age [6]. However, asthma heterogeneity and its multiple clinical manifestations reduce the accuracy of prediction.

2. Formal Problem Statement

Bronchial asthma, regardless of severity, is a chronic non-infectious inflammatory respiratory disease. The inflammatory process in airways causes airway hyperresponsiveness, bronchial obstruction and respiratory symptoms. Symptomatic manifestations of bronchial asthma, occurring more than twice a year, require a thorough examination of a child to exclude asthma. Due to under-diagnosis of asthma, the incorrect treatment is provided. Long-term ineffective therapy affects the formation of nonreversible blocking of bronchi as a result of bronchial wall remodeling. Asthma with a severe course calls a special attention and requires high drug dosage. [7, 8]. The multifactorial nature of asthma contributes to an uncontrolled course of the disease development [9, 10]. Predicting of severe asthma in children at risk will prevent exacerbations and improve a patient's quality of life. The identification of factors that make asthma more difficult to control will help to eliminate causative or precipitating factors [12, 13]. Using of linear regression models is a common approach to calculate a probability of a severe asthma onset or an uncontrolled course of disease development [14]. Regression models are used for preliminary estimation of bronchial asthma severity. As a rule, models contain from three to five regressors [15] where special mention should go to asthma in parents, atopic dermatitis, wheezes without a cold, specific Ig E, allergic rhinitis and a child's gender. To make a preliminary assessment of bronchial asthma severity, low-dimension models are considered because of the fact that the disease symptoms in a patient being studied are clearly caused by 5-7 factors. In most cases, these factors appear to be qualitative. At the same time, more than a hundred factors influence severity of bronchial asthma [16]. In most cases, the factors, on the one hand, are weakly related to each other, and on the other hand, they make approximately the same contribution to the formation of the explained quantity value [17]. The number of models required to describe possible situations corresponds to the number of ways where 5 factors can be selected out of a hundred of factors is determined by the value

$$\frac{100!}{5!(100-5)!} \sim 10^8, \tag{1}$$

Thus, despite a large number of publications devoted to predicting bronchial asthma severity, the proposed models can be used for a relatively small group of patients characterized by a certain set of regressors. To increase the scope of the model, it is possible to form some generalized factors that combine several regressors or to build models with a higher dimension. However, this approach has not become widely used in studies dealing with predicting bronchial asthma severity due to a significant computational complexity associated with a large number of multiple regression models building. As an alternative to these approaches, this paper proposes a technique for approximate regression models containing five factors building. The proposed methodology is based on the assumption that

a) the value of correlation coefficient $r_{x_m x_v}$ between the model's regressors X_m, X_v is small:

$$r_{x_m x_v} = \frac{K_{x_m x_v}}{\sqrt{D_{x_m} D_{x_v}}}, \ \left| r_{x_m x_v} \right| \to 0, \tag{2}$$

$$K_{x_m x_v} = \frac{\sum_{i=1}^{n} \left(X_{m\,i} - m_{x_m} \right)\left(X_{v\,i} - m_{x_v} \right)}{n}, \ m_{x_m} = \frac{\sum_{i=1}^{n} X_{m\,i}}{n}, \ D_{x_m} = \frac{\sum_{i=1}^{n} \left(X_{m\,i} - m_{x_m} \right)^2}{n}; \tag{3}$$

b) the value of correlation coefficient $\left| r_{x_m x_v} \right|$ is far smaller than the value of correlation coefficient $\left| r_{y\,x_m} \right|$ between the observed value Y and the model's regressor X_m:

$$\left| r_{y\,x_m} \right| >> \left| r_{x_m x_v} \right|, \qquad r_{y\,x_m} = \frac{K_{y\,x_m}}{\sqrt{D_y D_{x_m}}}, \tag{4}$$

$$K_{y x_m} = \frac{\sum_{i=1}^{n} \left(X_{m i} - m_{x_m}\right)\left(Y_i - m_y\right)}{n},\ m_y = \frac{\sum_{i=1}^{n} Y_i}{n},\ D_y = \frac{\sum_{i=1}^{n} \left(Y_i - m_y\right)^2}{n}; \quad (5)$$

These circumstances make it possible to build a multiple regression approximate model in the form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5, \quad (6)$$

$$\beta_m = \left(r_{y x_m} + 0\left(r_{x_m x_v}\right)\right)\frac{\sqrt{D_y}}{\sqrt{D_{x_m}}}; \quad (7)$$

$$\beta_0 = m_y - \sum_{m=1}^{5} \beta_m m_{x_m}, \quad (8)$$

where coefficients β_m of the regression model are determined to within the summands of the order of smallness $0\left(r_{x_m x_v}\right)$. The assumptions taken as the foundation in the method of approximate regression models building for bronchial asthma severity predicting are confirmed by the research results [18]. The simplicity of models building is the basis for widespread use of the proposed methodology in clinical trials

3. Literature Review

Linear regression models are the most common in predicting the risk of bronchial asthma in children. The paper [19] considers a linear regression model that analyzes the dependence of the observed value on five factors: wheezing after exercises; wheezing causing shortbreathing; coughing on exertion; atopic dermatitis and allergic sensitization. An eight-factor linear regression model (male gender, postterm delivery, parental education, parental inhalation medication, wheezing/dyspnea apart from colds, wheezing frequency, respiratory tract infections, and doctor's diagnosis of eczema (ever) and eczematous rash present) was used to analyze a group of 2877 children to determine the level of risk of asthma at a school age development [20]. A ten-factor model (gender, age, wheezes without a cold, wheezes rate, disruption of activity, shortbreathing, wheezes and cough caused by exercises and aeroallergens, atopic dermatitis, asthma or bronchitis in parents) was used in the two-staged study of risk of asthma at a school age and is described in the paper [21]. The regression model is based on a noninvasive predicting method and an increased number of factors. The absence of quantitative factors characterizing results of laboratory examinations in the model is a characteristic feature of these models. A linear regression model containing three quantitative (hospitalization, eczema and atopy in parents) and one qualitative factor (the positive and negative predictive value of specific Ig E to inhalant allergens) was studied in the paper [22]. This allowed us to increase a prognostic value of the model. In [23], a simplified technique that makes it possible to build a dependence of the observed value on quantitative factors was considered. A significant simplification was obtained as a result of using a combination of one-dimensional and three-dimensional models instead of the four-factor model. The uniqueness of the proposed approach is in reducing the computational complexity of the regression models building process. Another way to reduce the computational complexity is connected with the use of approximate methods of regression models [16, 17]. In these studies, a technique of one- and two-parameter approximate models building (TSLP, atopic dermatitis, allergic rhinitis, bronchial asthma in relatives of the second generation, pillow feather, domestic dust, severe) was proposed, as well as the use of their combinations to analyze the bronchial asthma severity in children predicting as a way to replace multifactor linear regression models. This paper expands the field of application of approximate methods for regression models building. The purpose of this paper is to develop a technique for the five-factor linear regression approximate models building and to substantiate the area of their use.

4. Analysis Data Preparation

To demonstrate the technique for five parametric regression approximate models building, we will use the results taken from a clinical study of the severity of bronchial asthma in children of Kharkov region, 2017 [18]. The study was conducted with respect for human rights and in accordance with international ethical requirements; it doesn't violate any scientific ethical standards and standards of biomedical research [24]. The group for asthma severity analyzing included 90 children at the age from 6 to 18. The structure of the group was as follows: the main subgroup contained 70 children with bronchial asthma and the control group with 20 healthy children. The data of the parents survey about the symptoms of patients, characteristic to bronchial asthma, as well as the history of the patients' diseases were analyzed. This information was the basis for the values of the qualitative factors used to build the regression model formation. The clinical features of the disease course were studied. The results of laboratory research were used to form the values of the model's quantitative factors.

During the analysis, 142 factors were under consideration. A detailed analysis of the factors is given in [16]. Using a set of clinical research data, the most significant factors in the study were identified to build a linear regression approximate model. For these factors, correlation coefficients $r_{x_m x_v}$ between the model regressors and correlation coefficients $r_{y x_m}$ between the regressor and the observed value are given in Table 1

Table 1
Correlation Coefficients Values

	Atopic dermatitis	Bronchial asthma in	Allergic rhinitis	Sheep wool	Domestic dust	Rabbit hair	Pillow feather	Bronchial asthma in father
Atopic dermatitis	-	-0.076	0.738	0.1346	0.158	0.1533	0.0181	0.1084
Bronchial asthma in	-0.076	-	0.1025	0.1483	0.2571	0.0069	0.2842	-0.087
Allergic rhinitis	0.738	0.1025	-	0.0058	0.1732	0.1107	0.1032	0.2704
Sheep wool	0.1346	0.1483	0.0058	-	0.1507	0.2658	0.3004	0.104
Domestic dust	0.158	0.2571	0.1732	0.1507	-	0.0424	0.1899	-0.040
Rabbit hair	0.1533	0.0069	0.1107	0.2658	0.0424	-	0.0727	0.3211
Pillow feather	0.0181	0.2842	0.1032	0.3004	0.1899	0.0727	-	-0.033
Bronchial asthma in father	0.1084	-0.087	0.2704	0.104	-0.040	0.3211	-0.033	-
Severe	0.3767	0.4157	0.3223	0.3373	0.3116	0.2236	0.3681	0.0309

Taking into account the selection criterion, we take six factors from Table 1 to build six proposed models

$$\left|r_{y x_m}\right| \to \max, \quad \left|r_{x_m x_v}\right| \to \min, \quad m,v = 1..M. \tag{9}$$

The numerical characteristics $m_{x_m}, m_y, D_{x_m}, D_y$ for the selected factors and the observed value (Severe) determining the mathematical expectation and variance are given in Table 2.

Table 2
Regressor numerical characteristics and the explained value

№	Regressor	m_{x_m}, m_y	D_{x_m}, D_y
1	Sheep wool	0.5217	0.4234
2	Rabbit hair	0.5652	0.7965
3	Bronchial asthma in relatives of second generation	0.0658	0.0614
4	Allergic rhinitis	0.4494	0.2474
5	Domestic dust	2.2319	1.280
6	Atopic dermatitis	0.0562	0.053
7	Severe	0.1124	0.0997

The numerical characteristics given in Table 1 and Table 2 were obtained as a result of clinical studies analysis, we used them to build the linear regression approximate models (6).

5. Linear Regression Approximate Model Building

To build a linear regression equation, we introduce some dimensionless parameters:

$$\eta = \frac{Y_i - m_y}{\sqrt{D_y}}, \quad \xi_m = \frac{X_{m\,i} - m_{x_m}}{\sqrt{D_{x_m}}}, \tag{10}$$

corresponding to the dimensionless value of the observed value and dimensionless values of the model regressors. Then the equation (6), taking into account the equality (7), can be written in its dimensionless form:

$$\eta = \alpha_1\xi_1 + \alpha_2\xi_2 + \alpha_3\xi_3 + \alpha_4\xi_4 + \alpha_5\xi_5, \tag{11}$$

$$\alpha_m = \beta_m \frac{\sqrt{D_{x_m}}}{\sqrt{D_y}}. \tag{12}$$

We use the least squares method [16] to determine the values of coefficients. From the condition for minimum of a mean square root error in the observed value predicting

$$\sum_{i=1}^{n}\left(\eta_i - \alpha_1\xi_{1i} + \alpha_2\xi_{2i} + \alpha_3\xi_{3i} + \alpha_4\xi_{4i} + \alpha_5\xi_{5i}\right)^2 \to \min \tag{13}$$

the system of equation follows:

$$\begin{cases}
\alpha_1\sum\limits_{i=1}^{n}\xi_{1i}\xi_{1i} + \alpha_2\sum\limits_{i=1}^{n}\xi_{1i}\xi_{2i} + \alpha_3\sum\limits_{i=1}^{n}\xi_{1i}\xi_{3i} + \alpha_4\sum\limits_{i=1}^{n}\xi_{1i}\xi_{4i} + \alpha_5\sum\limits_{i=1}^{n}\xi_{1i}\xi_{5i} = \sum\limits_{i=1}^{n}\xi_{1i}\eta_i, \\
\alpha_1\sum\limits_{i=1}^{n}\xi_{2i}\xi_{1i} + \alpha_2\sum\limits_{i=1}^{n}\xi_{2i}\xi_{2i} + \alpha_3\sum\limits_{i=1}^{n}\xi_{2i}\xi_{3i} + \alpha_4\sum\limits_{i=1}^{n}\xi_{2i}\xi_{4i} + \alpha_5\sum\limits_{i=1}^{n}\xi_{2i}\xi_{5i} = \sum\limits_{i=1}^{n}\xi_{2i}\eta_i, \\
\alpha_1\sum\limits_{i=1}^{n}\xi_{3i}\xi_{1i} + \alpha_2\sum\limits_{i=1}^{n}\xi_{3i}\xi_{2i} + \alpha_3\sum\limits_{i=1}^{n}\xi_{3i}\xi_{3i} + \alpha_4\sum\limits_{i=1}^{n}\xi_{3i}\xi_{4i} + \alpha_5\sum\limits_{i=1}^{n}\xi_{3i}\xi_{5i} = \sum\limits_{i=1}^{n}\xi_{3i}\eta_i, \\
\alpha_1\sum\limits_{i=1}^{n}\xi_{4i}\xi_{1i} + \alpha_2\sum\limits_{i=1}^{n}\xi_{4i}\xi_{2i} + \alpha_3\sum\limits_{i=1}^{n}\xi_{4i}\xi_{3i} + \alpha_4\sum\limits_{i=1}^{n}\xi_{4i}\xi_{4i} + \alpha_5\sum\limits_{i=1}^{n}\xi_{4i}\xi_{5i} = \sum\limits_{i=1}^{n}\xi_{4i}\eta_i, \\
\alpha_1\sum\limits_{i=1}^{n}\xi_{5i}\xi_{1i} + \alpha_2\sum\limits_{i=1}^{n}\xi_{5i}\xi_{2i} + \alpha_3\sum\limits_{i=1}^{n}\xi_{5i}\xi_{3i} + \alpha_4\sum\limits_{i=1}^{n}\xi_{5i}\xi_{4i} + \alpha_5\sum\limits_{i=1}^{n}\xi_{5i}\xi_{5i} = \sum\limits_{i=1}^{n}\xi_{5i}\eta_i,
\end{cases} \tag{14}$$

Its solution allows us to obtain the expressions for the values of coefficients α_m definition. Using dimensionless designations (10), as well as the form of expressions $K_{x_m x_v}$ (3), $K_{y x_m}$ (5), we obtain a formula for correlation coefficients through the introduced dimensionless parameters:

$$r_{x_m x_v} = \frac{K_{x_m x_v}}{\sqrt{D_{x_m} D_{x_v}}} = \frac{1}{n}\sum_{i=1}^{n} \xi_{mi}\xi_{vi}\ ,\quad r_{y\,x_m} = \frac{K_{y\,x_m}}{\sqrt{D_y D_{x_m}}} = \frac{1}{n}\sum_{i=1}^{n} \xi_{mi}\eta_i\ , \tag{15}$$

This allows us to represent the system of equations (14) in a simplified form:

$$\begin{cases} \alpha_1 + r_{x_1x_2}\alpha_2 + r_{x_1x_3}\alpha_3 + r_{x_1x_4}\alpha_4 + r_{x_1x_5}\alpha_5 = r_{y\,x_1}, \\ r_{x_1x_2}\alpha_1 + \alpha_2 + r_{x_3x_2}\alpha_3 + r_{x_4x_2}\alpha_4 + r_{x_5x_2}\alpha_5 = r_{y\,x_2}, \\ r_{x_1x_3}\alpha_1 + r_{x_2x_3}\alpha_2 + \alpha_3 + r_{x_4x_3}\alpha_4 + r_{x_5x_3}\alpha_5 = r_{y\,x_3}, \\ r_{x_1x_4}\alpha_1 + r_{x_2x_4}\alpha_2 + r_{x_3x_4}\alpha_3 + \alpha_4 + r_{x_5x_4}\alpha_5 = r_{y\,x_4}, \\ r_{x_1x_5}\alpha_1 + r_{x_2x_5}\alpha_2 + r_{x_3x_5}\alpha_3 + r_{x_4x_5}\alpha_4 + \alpha_5 = r_{y\,x_5}, \end{cases} \quad \begin{array}{l} r_{x_m x_v} = r_{x_v x_m}, \\ r_{y\,x_m} = r_{x_m y}\ , \end{array} \tag{16}$$

that has a solution regarding the unknown coefficients α_m:

$$\alpha_m = \frac{\Delta_m}{\Delta}\ , \tag{17}$$

$$\Delta = \begin{vmatrix} 1 & r_{x_1x_2} & r_{x_1x_3} & r_{x_1x_4} & r_{x_1x_5} \\ r_{x_1x_2} & 1 & r_{x_3x_2} & r_{x_4x_2} & r_{x_5x_2} \\ r_{x_1x_3} & r_{x_2x_3} & 1 & r_{x_4x_3} & r_{x_5x_3} \\ r_{x_1x_4} & r_{x_2x_4} & r_{x_3x_4} & 1 & r_{x_5x_4} \\ r_{x_1x_5} & r_{x_2x_5} & r_{x_3x_5} & r_{x_4x_5} & 1 \end{vmatrix} \quad \Delta_1 = \begin{vmatrix} r_{y\,x_1} & r_{x_1x_2} & r_{x_1x_3} & r_{x_1x_4} & r_{x_1x_5} \\ r_{y\,x_2} & 1 & r_{x_3x_2} & r_{x_4x_2} & r_{x_5x_2} \\ r_{y\,x_3} & r_{x_2x_3} & 1 & r_{x_4x_3} & r_{x_5x_3} \\ r_{y\,x_4} & r_{x_2x_4} & r_{x_3x_4} & 1 & r_{x_5x_4} \\ r_{y\,x_5} & r_{x_2x_5} & r_{x_3x_5} & r_{x_4x_5} & 1 \end{vmatrix}, \quad \Delta_2 = ... \tag{18}$$

If the conditions $\left|r_{y\,x_m}\right| >> \left|r_{x_m x_v}\right|$ (4) are met, coefficients α_m can be calculated using the approximate formula:

$$\alpha_m \approx \frac{r_{y\,x_m} - \sum_{v=1, v\neq m}^{M} r_{x_m x_v} r_{y\,x_v} + 0\left(r_{x_m x_v}{}^2\right)}{1 - \sum_{m}^{M}\sum_{v, v>m}^{M} r_{x_m x_v}{}^2 + 0\left(r_{x_m x_v}{}^3\right)} \approx r_{y\,x_m} - \sum_{v=1, v\neq m}^{M} r_{x_m x_v} r_{y\,x_v} + 0\left(r_{x_m x_v}{}^2\right) \tag{19}$$

To simplify the material presentation, we will consider the calculation of coefficients α_m to within the summands of the order of smallness $0\left(r_{x_m x_v}\right)$

$$\alpha_m \approx r_{y\,x_m} + 0\left(r_{x_m x_v}\right) \tag{20}$$

Taking into account (12), we express the value of coefficients β_m through the value of coefficients α_m

$$\beta_m = \alpha_m \frac{\sqrt{D_y}}{\sqrt{D_{x_m}}} \tag{21}$$

and substituting the values for the α_m (20) calculating, we obtain the linear regression equation in the form (6)-(8) to within the summands of the order of smallness $0\left(r_{x_m x_v}\right)$.

In the accepted approximation, the observed value depends only on the values of correlation coefficient $r_{x_m y}$ between the model regressors and the explained value. The linear regression equation can be used to preliminary estimate the severity of bronchial asthma. To improve the predicting accuracy for coefficients α_m calculating, the formula (19) that takes into account the values of correlation coefficients between the model regressors should be used.

6. Analysis of the Results

In this section, as an example, let us consider the building of five-factor linear regression approximate models to within the summands of the order of smallness $0\left(r_{x_m x_v}\right)$

$$\eta_a = \alpha_{1a}\xi_1 + \alpha_{2a}\xi_2 + \alpha_{3a}\xi_3 + \alpha_{4a}\xi_4 + \alpha_{5a}\xi_5 \quad (22)$$

The symbol "a" (approximate) in the model parameters means that to predict bronchial asthma severity when calculating the coefficients α_{ma} the approximate formula (20) was used. As the regressors, we will use the factors given in Table 2 and selected using the criterion (9). To calculate coefficients α_{ma} (20), we will use Table 1 showing the values of correlation coefficients between the model regressors and the explained value. The results of linear regression equations building for the selected factors are shown in Table 3.

Table 3
Linear Regression Models in Dimensionless Form

regressor	Missing element	exact dimensionless model (22) $\eta_e = \alpha_{1e}\xi_1 + \alpha_{2e}\xi_2 + \alpha_{3e}\xi_3 + \alpha_{4e}\xi_4 + \alpha_{5e}\xi_5$	approximate dimensionless model (23) $\eta_a = \alpha_{1a}\xi_1 + \alpha_{2a}\xi_2 + \alpha_{3a}\xi_3 + \alpha_{4a}\xi_4 + \alpha_{5a}\xi_5$	approximate model error (24) $\Delta\eta_a = \eta_a - \eta_e$
Sheepwool Rabbit hair Bronchial asthma Domestic dust Atopic dermatitis	Allergic rhinitis	$\eta_e = 0.185\xi_1 + 0.114\xi_2 + 0.381\xi_3 + 0.127\xi_5 + 0.343\xi_6$	$\eta_a = 0.337\xi_1 + 0.224\xi_2 + 0.416\xi_3 + 0.312\xi_5 + 0.377\xi_6$	$\Delta\eta_a = -0.152\xi_1 + (-0.110)\xi_2 - 0.035\xi_3 + (-0.185)\xi_5 - 0.034\xi_6$
Sheepwool Rabbit hair Allergic rhinitis Domestic dust Atopic dermatitis	Bronchial asthma	$\eta_e = 0.250\xi_1 + 0.773\xi_2 + 0.260\xi_4 + 0.181\xi_5 + 0.284\xi_6$	$\eta_a = 0.337\xi_1 + 0.224\xi_2 + 0.322\xi_4 + 0.312\xi_5 + 0.377\xi_6$	$\Delta\eta_a = -0.088\xi_1 + 0.549\xi_2 - 0.062\xi_4 + (-0.131)\xi_5 + 0.093\xi_6$
Sheep wool Bronchial asthma Allergic rhinitis Domestic dust Atopic dermatitis	Rabbit hair	$\eta_e = 0.223\xi_1 + 0.361\xi_3 + 0.223\xi_4 + 0.089\xi_5 + 0.342\xi_6$	$\eta_a = 0.337\xi_1 + 0.416\xi_3 + 0.322\xi_4 + 0.312\xi_5 + 0.377\xi_6$	$\Delta\eta_a = -0.108\xi_1 + (-0.05)\xi_3 - 0.08\xi_4 + (-0.223)\xi_5 + 0.032\xi_6$
Rabbit hair Bronchial asthma Allergic rhinitis Domestic dust Atopic dermatitis	Sheep wool	$\eta_e = 0.138\xi_2 + 0.390\xi_3 + 0.222\xi_4 + 0.112\xi_5 + 0.351\xi_6$	$\eta_a = 0.224\xi_2 + 0.416\xi_3 + 0.322\xi_4 + 0.312\xi_5 + 0.377\xi_6$	$\Delta\eta_a = -0.086\xi_2 + (-0.026)\xi_3 - 0.1\xi_4 + (-0.2)\xi_5 - 0.026\xi_6$
Sheep wool Rabbit hair Bronchial asthma Allergic rhinitis Atopic dermatitis	Domestic dust	$\eta_e = 0.210\xi_1 + 0.085\xi_2 + 0.385\xi_3 + 0.247\xi_4 + 0.347\xi_6$	$\eta_a = 0.337\xi_1 + 0.224\xi_2 + 0.416\xi_3 + 0.322\xi_4 + 0.377\xi_6$	$\Delta\eta_a = -0.128\xi_1 + (-0.139)\xi_2 - 0.031\xi_3 + (-0.076)\xi_4 - 0.03\xi_6$
Sheepwool Rabbit hair Bronchial asthma Allergic rhinitis Domestic dust	Atopic dermatis	$\eta_e = 0.233\xi_1 + 0.127\xi_2 + 0.317\xi_3 + 0.250\xi_4 + 0.146\xi_5$	$\eta_a = 0.337\xi_1 + 0.224\xi_2 + 0.416\xi_3 + 0.322\xi_4 + 0.312\xi_5$	$\Delta\eta_a = -0.104\xi_1 + (-0.098)\xi_2 - 0.099\xi_3 + (-0.073)\xi_4 - 0.165\xi_5$

We inject an accurate linear regression model

$$\eta_e = \alpha_{1e}\xi_1 + \alpha_{2e}\xi_2 + \alpha_{3e}\xi_3 + \alpha_{4e}\xi_4 + \alpha_{5e}\xi_5 \quad (23)$$

where, in parameters, the symbol "e" (exact) means that coefficients α_{1e} are calculated in accordance with the formula (17) without the introduction of any assumptions concerning the values of correlation

coefficients $r_{x_m x_v}$ smallness between the model regressors. The results of linear regression equations building are also given in Table 3.

We subtract the equation (23) from the equation (22) and obtain an expression for the approximation error estimation

$$\Delta\eta_a = \eta_a - \eta_e = \sum_{m=1}^{5} (\alpha_{ma} - \alpha_{me})\xi_m \qquad (24)$$

associated with replacing the exact model with an approximate one. Results of a comparative assessment of the models (22) and (23) are given in Table 3

Table 4

Results of Analysis of the Linear Regression Approximate Model Use

regressor	missing element	exact model (26)	approximate model (25)	approximate model error $\Delta Y_a = (Y_e - Y_a)$
Sheep wool Rabbit hair Bronchial asthma Domestic dust Atopic dermatitis	Allergic rhinitis	$Y_e = -0.095 + 0.09X_1 + 0.04X_2 + 0.485X_3 + 0.035X_4 + 0.471X_5$	$Y_a = -0.227 + 0.164X_1 + 0.079X_2 + 0.53X_3 + 0.066X_4 + 0.501X_5$	$\Delta Y_a = 0.133 - 0.074X_1 + (-0.039)X_2 - 0.045X_3 + (-0.03)X_4 - 0.3X_5$
Sheep wool Rabbit hair Allergic rhinitis Domestic dust Atopic dermatitis	Bronch. asthma	$Y_e = -0.175 + 0.121X_1 + 0.027X_2 + 0.165X_4 + 0.051X_5 + 0.389X_6$	$Y_a = -0.284 + 0.164X_1 + 0.079X_2 + 0.205X_4 + 0.066X_5 + 0.501X_6$	$\Delta Y_a = 0.109 - 0.942X_1 + (-0.052)X_2 - 0.04X_4 + (-0.015)X_5 - 0.113X_6$
Sheep wool Bronchial asthma Allergic rhinitis Domestic dust Atopic dermatitis	Rabbit hair	$Y_e = -0.126 + 0.108X_1 + 0.460X_3 + 0.155X_4 + 0.025X_5 + 0.469X_6$	$Y_a = -0.275 + 0.164X_1 + 0.530X_3 + 0.205X_4 + 0.066X_5 + 0.501X_6$	$\Delta Y_a = 0.149 - 0.056X_1 + (-0.07)X_3 - 0.05X_4 + (-0.041)X_5 - 0.032X_6$
Rabbit hair Bronchial asthma Allergic rhinitis Domestic dust Atopic dermatitis	Sheep wool	$Y_e = -0.108 + 0.05X_2 + 0.497X_3 + 0.141X_4 + 0.031X_5 + 0.482X_6$	$Y_a = -0.234 + 0.08X_2 + 0.53X_3 + 0.205X_4 + 0.066X_5 + 0.501X_6$	$\Delta Y_a = 0.126 - 0.03X_2 + (-0.33)X_3 - 0.064X_4 + (-0.035)X_5 - 0.02X_6$
Sheep wool Rabbit hair Bronchial asthma Allergic rhinitis Atopic dermatitis	Domes-tic dust	$Y_e = -0.087 + 0.102X_1 + 0.03X_2 + 0.491X_3 + 0.157X_4 + 0.475X_6$	$Y_a = -0.173 + 0.164X_1 + 0.079X_2 + 0.53X_3 + 0.205X_4 + 0.501X_6$	$\Delta Y_a = 0.086 - 0.062X_1 + (-0.049)X_2 - 0.039X_3 + (-0.048)X_4 - 0.026X_6$
Sheep wool Rabbit hair Bronchial asthma Allergic rhinitis Domestic dust	Atopic dermat.	$Y_e = -0.161 + 0.113X_1 + 0.044X_2 + 0.404X_3 + 0.158X_4 + 0.041X_5$	$Y_a = -0.291 + 0.164X_1 + 0.079X_2 + 0.53X_3 + 0.205X_4 + 0.066X_5$	$\Delta Y_a = 0.13 - 0.05X_1 + (-0.035)X_2 - 0.126X_3 + (-0.046)X_4 - 0.025X_5$

Using relation (21) between the parameters α_m and β_m, we proceed from the dimensionless model (22) to its analogue (6), where, while calculating coefficients α_m and β_m, the expression (20) was used

$$Y_a = \beta_{0a} + \beta_{1a}X_1 + \beta_{2a}X_2 + \beta_{3a}X_3 + \beta_{4a}X_4 + \beta_{5a}X_5 \qquad (25)$$

The analysis of the research results given in Table 3 demonstrates a satisfactory accuracy of the linear regression approximate model. The error of replacing an exact model with an approximate one corresponds to the expected value of the error, defined as $0\left(r_{x_m x_v}\right)$.

Results of proceeding from the linear regression dimensionless approximate model (22) to the model (25) are given in Table 4. Also, let us inject an accurate linear regression model

$$Y_e = \beta_{0e} + \beta_{1e} X_1 + \beta_{2e} X_2 + \beta_{3e} X_3 + \beta_{4e} X_4 + \beta_{5e} X_5 \quad (26)$$

that corresponds to the model (23). Results of proceeding from the model (23) to the model (26) as well as an approximation error of this proceeding $\Delta Y_a = \left(Y_e - Y_a\right)$ are given in Table 4.

This approximation is used to demonstrate the technique of approximate models building for preliminary prediction of bronchial asthma severity. The value of the error that occurs when replacing an accurate linear regression model with an approximate model can be estimated using this formula

$$0\left(r_{x_m x_v}\right) \approx \sum_{v=1, v \neq m}^{M} r_{x_m x_v} r_{y x_v} \quad (27)$$

In more detailed studies, a linear regression approximate model where the coefficients α_m and β_m are determined to within the summands of the order of smallness $0\left(r_{x_m x_v}{}^2\right)$, should be used. The error of replacing an exact model with an approximate one significantly depends on the value of a correlation coefficient between model regressors. Therefore, analysis of the approximate model area of application is of utmost importance. The cases when the parameters of the model satisfy both condition (9) and the requirement $\left|r_{y x_m}\right| >> \left|r_{x_m x_v}\right|$ are considered as the most successful use of approximate models.

To calculate the coefficients of the regression model, specialized software was developed using an object-oriented programming language Java (java 1.8; jdk 1.8.0_162). Mathematical operations on matrices were carried out using the open source mathematical package org.apache.commons commons-math3 (version 3.2) [25], available under license: The Apache Software License, Version 2.0. Computing resources are involved in the calculation: processor Intel®Core™ i7-4790CPU 3.6GHz; RAM 32GB; OS: Windows 10, 64-bit.

7. Conclusion

In this paper, the application of technique of a five-parameter approximate model building to predict bronchial asthma severity is discussed. The fact that, when predicting bronchial asthma severity, the observed value depends on a large number of factors usually weakly related to each other is the ground of the proposed technique. Depending on the region, age, living conditions, more than a hundred of factors causing the disease can be revealed. A preliminary examination of a patient usually reveals from 5 to 7 factors characterizing the disease severity. This explains the existence of publications discussing the building of a regression model with 3-5 factors. In each case, the specified set of factors is different. This requires a huge number of linear regression models, the building of which is associated with significant computational difficulties. The developed technique allows us to simplify the process of five-factor linear regression models building significantly. The fact that this technique doesn't require any massive computational resources to build linear regression multiparameter models is one of its important advantages. It is a good way to preliminary analyze the disease severity. In this paper, not only the technique of linear regression approximate model building but also the area of its application as well as an error of proceeding from an accurate model to an approximate one are analyzed. The presence of strong correlations between the model regressors has a significant effect on the specified error. Two approximations that can be used to build linear regression models are discussed in details. As a shallow analysis, an approximation without correlations between the model regressors can be used. For a detailed analysis, an approximation with linear dependence of model coefficients on correlation coefficients between model regressors taken into account while calculating these model coefficients should be used. Approximations with nonlinear dependences taken into account when calculating the model coefficients are a prospect for further studies. Building of linear regression approximate models

containing 10-15 regressors as well as the analysis of the proceeding error from an exact model to an approximate one dependence on the number of regressors in the model is of particular practical interest.

This article discusses the type of linear multivariate regression model. The choice of the type of model and the number of factors in the model is an urgent issue in predicting the severity of bronchial asthma disease. This circumstance determines the prospects for further research: 1) comparative analysis of the accuracy of predicting the severity of bronchial asthma disease depending on the type of regression model with the same number of model regressors; 2) comparative analysis of the accuracy of predicting the severity of bronchial asthma disease depending on the number of factors used in the model; 3) selection of a criterion for assessing the quality of prediction.

8. References

[1] WHO. Asthma/ WHO, 2020). https://www.who.int/news-room/fact-sheets/detail/asthma.

[2] M. Fajt, S. Wenzel, Asthma phenotypes and the use of biologic medications in asthma and allergic disease: The next steps toward personalized care,. Journal of Allergy and Clinical Immunology, vol. 135, 2015, pp. 299–310. https://doi.org/10.1016/j.jaci.2014.12.1871.

[3] GINA, 2020. https://ginasthma.org/wp-content/uploads/2020/04/GINA-2020-full-report_-final-_wms.pdf.

[4] R. Zeiger, M. Schatz, A. Dalal, L. Qian, W. Chen, E. Ngor, et al., Utilization and costs of severe uncontrolled asthma in a managed-care setting, J Allergy Clin Immunol Pract, 2016, 4:120–9.e3. doi.10.1016 / j.jaip.2015.08.003.

[5] L. Fleming, C. Murray, A. Bansal, S. Hashimoto, H. Bisgaard, A. Bush. and et al, The burden of severe asthma in childhood and adolescence: results from the pediatric U-BIOPRED cohorts, European Respiratory Journal, vol. 46, 2015, pp. 1322–1333. https://doi.org/ 10.1183/13993003.00780-2015.

[6] G. Luo, F. Nkoy, B. Stone, D. Schmick, M. Johnson, A systematic review of predictive models for asthma development in children, BMC Med Inform Decision Making, 15:99, 2015.

[7] E. Barsky, L. Giancola, S. Baxi, J. Gaffin, A practical approach to severe asthma in children, Ann Am Thorac Soc, vol. 15, 2018, pp. 399–408. https://doi.org/10.1513/AnnalsATS.201708-637FR.

[8] W. Phipatanakul, D. Mauger, R. Sorkness, J. Gaffin, F. Holguin, P. Woodruff, et al, Severe Asthma Research Program. Effects of age and disease severity on systemic corticosteroid responses in asthma, Am J Respir Crit Care Med, vol. 195(11), 2017, pp. 1439–1448. https://doi.org/ 10.1164/rccm.201607-1453OC.

[9] M. Zaniboni, L. Samorano, R. Orfali, V. Aoki, Skin barrier in atopic dermatitis: beyond filaggrin, An Bras. Dermatol, vol. 91, 2016, pp. 472–478. https://doi.org/10.1590/abd1806-4841.201644.

[10] I. Pavord, N. Hanania, Controversies in allergy: should severe asthma with eosinophilic phenotype always be treated with anti-IL-5 therapies, The Journal of Allergy and Clinical Immunology: In Practice, vol. 7, 2019, pp.1430–1436. doi:10.1016/j.jaip.2019.03.010.

[11] T. Guilbert, L. Bacharier, A. Fitzpatrickc, Severe asthma in children, J Allergy Clin Immunol Pract, vol. 2(5), 2014, pp. 489–500. doi: 10.1016/j.jaip.2014.06.022.

[12] P. Amin, L. Levin, T. Epstein, P. Ryan, G. LeMasters, G. Khurana, et al, Optimum predictors of childhood asthma: persistent wheeze or the asthma predictive index? J Allergy Clin Immunol Pract, vol. 2(6), 2014, pp. 709–715. https://doi: 10.1016/j.jaip.2014.08.009.

[13] M. Morais-Almeida, A. Gaspar, G. Pires, S. Prates, J. Rosado-Pinto, Risk factors for asthma symptoms at school age: an 8-year prospective study, Allergy Asthma Proc, 28(2), 2007, 183-189. doi: 10.2500/aap.2007.28.2953.

[14] S. van der Werff, R. Junco, R. Reyneveld, M. Heymans, M. Ponce, M. Gorbea, et al, Prediction of asthma by common risk factors: a follow-up study in Cuban schoolchildren, J Investig Allergol Clin Immunol, vol. 23(6), 2013, pp. 415–420. http://www.jiaci.org/issues/vol23issue6/6.pdf.

[15] D. Kothalawala, L. Kadalayil, V. Weiss, M. Aref Kyyaly, S. Hasan Arshad, J. Holloway, et al, Prediction models for childhood asthma: a systematic review, Pediatr Allergy Immunol, vol. 31, 2020, pp. 616–627. https://doi.org/10.1111/pai.13247.

[16] O. Pihnastyi, O. Kozhyna, Use of the complex of models of regression for analysis of the factors that determine the severity of bronchial asthma, International Medicine, vol. 2(2), 2020, pp. 107– 118. http://dx.doi.org/10.5455/im.74961.
[17] O. Pihnastyi, O. Kozhyna, Methods for constructing estimated two-factor linear regression models for diagnosing the severity of bronchial asthma in children, Innovare Journal of Medical Sciences, vol. 9(1), 2021, pp. 23–30. doi: https://doi.org/10.22159/ijms.2021.v9i1.40408.
[18] O. Kozhyna, O. Pihnastyi, Covariance coefficients factors from a clinical study of the severity of bronchial asthma in children of the Kharkov region, 2017, Mendeley Data, 1, 2019.
[19] R. Wang, A. Simpson, A. Custovic, P. Foden, D. Belgrave, C. Murray, Individual risk assessment tool for school-age asthma prediction in UK birth cohort, Clin Exp Allergy, , vol. 49(3), 2019, pp. 292-298.
[20] E. G. Hafkamp-de, H. F. Lingsma, D. Caudri, D. Levie, A. Wijga, G. H. Koppelman, et al, Predicting asthma in preschool children with asthma-like symptoms: validating and updating the PIAMA risk score, J Allergy Clin Immunol, vol. 132(6), 2013, pp.1303-1310. doi: 10.1016/j.jaci.2013.07.007.
[21] A. Pescatore, C. Dogaru, L. Duembgen, M. Silverman, E. Gaillard, B. Spycher, et al, A simple asthma prediction tool for preschool children with wheeze or cough, J Allergy Clin Immunol, vol. 133(1), 2014, pp. 111-8.e1-13. doi: 10.1016/j.jaci.2013.06.002.
[22] N. Boersma, R. Meijneke, J. Kelder, C. van der Ent, W. Balemans, Sensitization predicts asthma development among wheezing toddlers in secondary healthcare, Pediatr Pulmonol, vol. 52(6), pp. 729-736, 2017.
[23] R. Kurukulaaratchy, S. Matthews, S. Holgate, S. Arshad, Predicting persistent disease among children who wheeze during early life, Eur Respir J, vol. 22(5), 2003, pp. 767-771.
[24] UNESCO.ORG, Universal Declaration on Bioethics and Human Rights, 2005. http://portal.unesco.org/en/ev.php. 2005.
[25] Apache Commons Math. http://commons.apache.org/proper/commons-math/