

UDC 657:004.94

JEL: J 01; O21; O35; M50

REGRESSION MODELS FOR PROJECTING LIFE EXPECTANCY IN ADMINISTRATIVE-TERRITORIAL FORMATIONS: CONSTRUCTION AND QUALITY ESTIMATION

©*Khubaev G., Dr. habil., Rostov State University of Economics (RINH),
Rostov-on-Don, Russia, gkhubaev@mail.ru*

Abstract. The article presents statistically significant regression models ($R^2_{\text{corr}} > 0,9$; $F_{\text{cr}} \gg 100$; response function — life expectancy). The analysis of the abnormal observations in the baseline information has been carried out according to an original approach.

Keywords: life expectancy, correlation coefficients, regression models, random errors.

Formulation of the problem

It is known that the summary indicator of the level of human development in a country (so called *life quality* or *the level of life*), Human Development Index (HDI) is annually calculated by the experts of the United Nations Development Programme. The Index was first introduced in the United Nations Human Development Report in 1990. HDI indicates the country's achievements in the aspects of life expectancy, education and average income (per capita income) according to the parameters of life expectancy at birth (for health and longevity), knowledge (the adult literacy rate and the gross enrollment ratio) and the standard of living (indicated by the natural logarithm of gross domestic product (GDP) per capita at purchasing power parity (PPP).

However, the information on the factors that would have the *statistically significant* influence on the Human Development Index and the country achievements rate has not been confirmed by any *reliable* quantitative results of cross-sectional studies. For instance, what is the *statistically significant* coherence between the longevity indicator and the combination of other factors involved in statistical calculations?

The article presents the original research results on the building statistically significant regressive models to project life expectancy in Administrative–Territorial Formations with revealing and estimating *the statistical significance* of a score of factors influencing life expectancy.

The Federal State Statistic Service official website and the Statistical Abstracts provided us with the socio-economic indicators of the Russian Federation Administrative-Territorial Formations, the basic data for the research [1]. The seven-year period data was investigated (2010–2016). According to the fact that the Statistical Abstract 2017 provides no information on 2 of 135 indicators for 2016 (*Gross Regional Product* and *Per Capita Gross Regional Product*), the figures of 2015 were taken for the preliminary calculations. The results were cross-checked with the data of the previous years and 2016 in case when the two parameters could be omitted.

1. Estimating the statistical significance of the determining factors

The *meaningful analysis* allowed to select some tens of indicators from the amount of the socio-economic parameters characterizing the administrative divisions of the Russian Federation. The selected indicators were studied to reveal the correlation relationship to the *Life Expectancy at Birth Index*. The correlation was checked according to R.A. Fisher: *t*-distribution with the degrees of freedom ($n-2$), $n=85$. The number of the indicators with the *rejected null hypothesis* exceeded 30.

On the next stage we investigated the possibility to create the functional relationship between the *Life Expectancy at Birth Index* (response function) and the indicators statistically related to the parameter (the indicators with statistically significant correlation coefficients were used as

independent variables). Conducting the experiments to find the model with the best statistical criteria, the regression equations were created with excitation variables of different combinations of factors-indicators. In fact, it caused a definite problem — even 20 elements could produce a great number of 2-, 3-element combinations.

However, the most of the created regressive models having the statistically significant b -coefficients failed to approximate the baseline statistical information well enough ($R^2 < 0.5$). For example, the regression models with various combinations of the indexes such as *Students Attending Bachelor, Specialist, Master Degree Programs, The Number of Enrolled Students per 100000 of the Population, The Number of Hospital Beds per 10000 of the Population, The Number of Nurses per 10000 of the Population, The Number of Doctors per 10000 of the Population* etc., including Education Index (*Students Attending Primary, Basic and General Secondary Education Programs + Students Attending Skilled Worker Programs*)* k_1 + (*Students Attending Mid-Career Professional Programs*)* k_2 + (*Students Attending Bachelor, Specialist, Master Degree Programs*)* k_3 ; ($k_1+k_2+k_3=1$), even with statistically significant b -coefficients revealed low coefficients of R^2 - and F -criterion.

Therefore, the subject ranking universal algorithm [2–3] was used to create projection models serving to reveal the relations between the objects; form the sub-groups with approximately the same indicators value and analyze the reasons for the sub-group formation process; expand the possibilities for the high-level system to make an optimal choice of the object, providing it with the information on the objects with the leading position and their criteria; provide each object with the possibility to analyze the potential of improving its activity, to offer a preliminary esteem of its competitive position; calculate statistical characteristics (mathematical expectation, variance, coefficient of variation, midpoint, asymmetric, kurtosis), tables and histograms of the distribution (the distribution of Administrative–Territorial Formations according to the numeral values of the indicators); estimate the dynamics of the successfully functioning objects; provide the relevant estimation of the high-level system.

The further analyses and creating of the regressive models were based on the background data presented in relative units.

Finally, the statistically significant regression models, providing a well-described basic data, with a combination of the factors, which appearance in the regression equation were meaningfully approved had been obtained as the result.

The regression model without an intercept term:

$$Y = b_1 * X_1 + b_2 * X_2 + b_3 * X_3,$$

Where Y is the relative value of the *Life Expectancy Index for an Administrative-Territorial Formation* (relating to the mean value of the indicator for the population of the Russian Federation administrative divisions); X_1 — relative level of education in an Administrative-Territorial Formation (the aggregate relative value of the number of students); X_2 — relative number of medical personnel (summarized relative rates of *The Number of Nurses per 10000 of the Population* and *The Number of Doctors per 10000 of the Population* indexes); X_3 — relative value of *The Number of Hospital Beds per 10000 of the Population* index.

The calculations resulted in the following values of the statistical indexes of the model quality:

For the models with 3-factor independent variables (per year)

*the factor values are presented in relative units

2015

b -coefficient values:

$b_1=1.115; b_2=23.72; b_3=17.53;$
 $b_1/\sigma_{b_1}=3.64; b_2/\sigma_{b_2}=5.88; b_3/\sigma_{b_3}=2.27;$
 $R^2=0.97; R^2_{\text{corr}}=0.967; F_{\text{cr}}=1323.1.$

2016

$b_1=0.0158; b_2=0.325; b_3=0.27;$
 $b_1/\sigma_{b_1}=3.785; b_2/\sigma_{b_2}=5.89; b_3/\sigma_{b_3}=2.52;$
 $R^2=0.98; R^2_{\text{corr}}=0.968; F_{\text{cr}}=1375.8.$

2014

$b_1=0.0138; b_2=0.36; b_3=0.205;$
 $b_1/\sigma_{b_1}=3.056; b_2/\sigma_{b_2}=6.03; b_3/\sigma_{b_3}=1.8;$
 $R^2=0.978; R^2_{\text{corr}}=0.966; F_{\text{cr}}=1258.1.$

*the absolute scale was used to calculate the factors values:

$b_1=0.0197; b_2=0.31; b_3=0.18;$
 $b_1/\sigma_{b_1}=4.77; b_2/\sigma_{b_2}=6.38; b_3/\sigma_{b_3}=1.8;$
 $R^2=0.98; R^2_{\text{corr}}=0.97; F_{\text{cr}}=1487.5.$

2) For the models with 3-factor independent variables

2016

*the factor values are presented in relative units

X_1 — relative education level, i. e. the relative number of students is divided by the relative population figures; X_2 — the sum of the relative number of doctors and the relative number of nurses per 10000 of the population in an Administrative-Territorial Formation.

$b_1=0.103; b_2=0.283;$
 $b_1/\sigma_{b_1}=5.14; b_2/\sigma_{b_2}=6.95;$
 $R^2=0.98; R^2_{\text{corr}}=0.97; F_{\text{cr}}=2250.4.$

The numeric values of the regression coefficients estimation and the statistical criterions of the regressive models based on the data of 2014 and 2015 do not differ in particular from the data of 2016.

*the absolute scale was used to calculate the factors values:

2016

$b_1=260.45; b_2=0.192;$
 $b_1/\sigma_{b_1}=6.93; b_2/\sigma_{b_2}=5.32;$
 $R^2=0.98; R^2_{\text{corr}}=0.977; F_{\text{cr}}=2716.$

So, how *the population and the individuals in charge of making decisions* (IMD) can benefit from the offered models? What is *the particular social* usefulness of these models?

In our opinion, the *statistically significant* regression equations presented in the paper describe the basic data (as it has appeared, *relevant enough*) with *insignificant errors* and reveal their usefulness in the following:

1) The individuals in charge of making decisions (IMD) on the different administrative levels are now provided with *an approved set of determining factors* so they could make optimization calculations to choose the most appropriate option for minimizing the reserves invested in the longevity programs: to support the increasing rates of the medical personnel and their professional level improvement or/and invest in medical facilities; or allocate funds to improve the *education level* of the population and *reduce the resource intensity* [4–5] in order to lower the cost of *the consumption basket* (to provide the customers with possibility to buy more, better and healthier products) etc.;

2) The society with the increased education level (due to the measures taken by the authorities, as mentioned above) would obtain the possibility *to realize that the life expectancy of each citizen's children and grand-children would be obviously increasing.*

We would like to draw the attention to the point that, having identified the list of the basic indicators presumably related to the life level and the life quality characteristics (due to the expert or quantitative analysis [6]), it becomes possible to form a statistically significant set of determining factors. Managing the factors could ensure a better quality of life in Russia or any other country as well.

2. Substantial rationale for the determining factors selection

The calculations made according to the baseline information presented on the Russian Federation Federal State Statistics Service official site resulted in the *statistically significant* models which correlate the *Expected Life Expectancy of the Administrative-Territorial Formations Index* with the indicators *X1* (the relative level of education in an Administrative-Territorial Formation); *X2* (the relative number of medical personnel); *X3* (the relative number of hospital beds).

What is more, the factors-indicators included into the model (actually, their appearance in the model is statistically defensible) have proved their statistical significance as well as their participation in the model seems to be easy to substantiate.

Obviously, a higher education level rate would greatly contribute to the wider public knowledge on the principles of healthy living and the latest achievements of science and medicine. Continuing in this direction, the inclusion of a new indicator (showing the relation of the student number to the population number in the administrative division) in the independent variables set becomes rationale.

Even the appearance of the *Hospital Bed Number per 10 000 of the Population Index* in the model could be easily explained, though it may seem not to be a simple case. The situation becomes clear as soon as we take into consideration the fact that the greater number of hospital beds indicates the greater number of medical personnel which results in increasing professional competitiveness and greater opportunities for professional improvement which attract outstanding specialists to the area.

3. Revealing abnormal observations in the baseline information

The research process involved a great amount of the baseline statistical information: 85 administrative divisions, 135 parameters characterizing their activity which provide more than 10 000 units of numerical data per year. It is obvious that such amount of figures would contain abnormalities, errors and outliers. How could the abnormal observations be distinguished and extracted from the baseline data? The abnormalities might negatively influence the calculation results, the conclusions and the quality of taken decisions.

Therefore, we used the abnormal observation detection method to find and extract the abnormalities from the Federal State Statistics Service database used as the baseline information to create regression models. According to the method the correlation matrixes and regression equations were created.

The analysis of the correlation matrixes calculated for each group of the identified factors-indicators (for a three-year period) alongside the analysis of residuals in two variants (absolute and normalized) revealed abnormal observations in the baseline data of 2 administrative divisions of the Russian Federation (the average correlation coefficient values of the divisions differ on the order from other divisions' average values). Anyway, the abnormality of 2 divisions had no particular influence on the quality of the created regressive models.

Conclusion

The conclusions that can be drawn are the following:

1. To our knowledge, this is the *first report* of the regression models created to project the live expectancy in the administrative-territorial formations that prove to be of a better quality (estimated according the standard criteria of statistical significance $R^2_{corr} > 0,9$; $F_{cr} \gg 100$) with predictive qualities improved by “crossed” estimates (based on the different years data); using open official statistical data with the simultaneous analysis of the abnormal observations in the baseline information.

2. The set of the independent variables of the created regression models was expanded with new statistically significant factors-indicators (the most cases show b-coefficient relation $b_i/\sigma_{b_i} \gg 2$) original in their structure and substantially approved.

3. The research provides the possibility to make optimization calculations based on the subset of the determining factors identified within the investigation process; the calculations would result in minimizing the reserve inverted in the projects aimed to increase the Life Expectancy Index for an Administrative-Territorial Formation.

Acknowledgments

The author would like to thank Mr. Daniil Korotin and Mr. Dmitriy Sidorenko for their helpful assistance in preparing the baseline information for the paper that allowed to reduce greatly the calculation efforts required.

The article is prepared based on the results of studies funded by the Russian Foundation for Basic Research (RFBR) — Project 18-010-00806/18.

References:

1. Regions of Russia. (2017). The main characteristics of the subjects of the Russian Federation. Statistical collection. Moscow, URL: <https://goo.gl/nSRi5r>. (in Russian).
2. Khubaev, G. N. (2018). Ranking of objects by a lot of quantitative indicators: universal algorithm. *RISK: Resources, information, supply, competition*, (1), 213-217. (in Russian).
3. Khubaev, G. N. (2018). Quality of life of the population of administrative-territorial entities: the method of express analysis. System analysis in design and management (SAEC-2018). *Proceedings of the XXII International Scientific and Practical Conference (St. Petersburg, SPbPU Peter the Great, May 22-24, 2018). V. 2. St. Petersburg, Polytechnic. Univ., 139-146.* (in Russian).
4. Khubaev, G., Kalugyan, K., Rodina, O., Shcherbakov, S. & Shirobokova, S. (2016). Universal methodical and tool support rapid assessment and optimization of resource-intensive goods and services. *Bulletin of Science and Practice*, (12), 286-299. (in Russian).
5. Khubaev, G. N., Shcherbakov, S. M., Shirobokova, S. N. (2016). Algorithms and software for rapid assessment of the resource intensity of goods and services. *Commonwealth (Scientific Russian-Chinese Journal)*, (9), 160-167. (in Russian).
6. Khubaev, G. (2017). Expert review: method of intuitively agreed choice. *In: Proceeding 5th International Conference “Economy modernization: new challenges and innovative practice” (November 12, 2017, Sheffield, UK)*, 65-80.
7. Khubaev, G. N. (2014). The way to identify errors in large numbers of numerical information. *Questions of statistics*, (10), 20-25. (in Russian).

УДК 657:004.94
JEL: J 01; O21; O35; M50

РЕГРЕССИОННЫЕ МОДЕЛИ ДЛЯ ПРОГНОЗИРОВАНИЯ ПРОДОЛЖИТЕЛЬНОСТИ ЖИЗНИ НАСЕЛЕНИЯ АДМИНИСТРАТИВНО- ТЕРРИТОРИАЛЬНЫХ ОБРАЗОВАНИЙ: ПОСТРОЕНИЕ И ОЦЕНКА КАЧЕСТВА

©Хубаев Г. Н., д-р экон. наук, Ростовский государственный экономический университет
(РИНХ), г. Ростов-на-Дону, Россия, gkhubaev@mail.ru

Аннотация. Построены статистически значимые регрессионные модели ($R^2_{\text{скорр}} > 0,9$; $F_{\text{кр}} \gg 100$; функция отклика — ожидаемая продолжительность жизни населения). С использованием оригинальной методики выполнена проверка наличия аномальных наблюдений в массиве исходной статистической информации, содержащем десятки тысяч числовых данных.

Ключевые слова: продолжительность жизни, коэффициенты корреляции, регрессионные модели, случайные ошибки.

Постановка задачи

Известно, что суммарный показатель уровня развития человека в стране (так называемого «качества жизни» или «уровня жизни») — Индекс развития человеческого потенциала (Human Development Index, HDI) — ежегодно рассчитывается экспертами Программы развития ООН (ПРООН) и с 1990 года используется ООН в ежегодном отчете по развитию человеческого потенциала. Индекс измеряет достижения страны с точки зрения продолжительности жизни, получения образования и фактического дохода, по трем основным направлениям: здоровье и долголетие, измеряемые показателем *ожидаемой продолжительности жизни при рождении*, доступ к образованию, измеряемый *уровнем грамотности взрослого населения* и совокупным валовым коэффициентом охвата образованием, достойный уровень жизни, измеряемый *величиной валового внутреннего продукта* (ВВП) на душу населения в долларах США по паритету покупательной способности (ППС).

Однако в настоящее время *отсутствуют достоверные* количественные, подтвержденные результатами перекрестных исследований *сведения* о том, какие факторы реально, *статистически значимо* влияют на Индекс развития человеческого потенциала, на достижения страны, какова, например, *статистически значимая* корреляционная связь между продолжительностью жизни населения и совокупностью различных, учитываемых статистическими органами факторов.

В статье представлены оригинальные результаты исследований, связанных с *построением* статистически значимых регрессионных моделей для прогнозирования ожидаемой продолжительности жизни населения административно-территориальных образований, с *выявлением и оценкой статистической значимости* влияния нескольких десятков факторов на ожидаемую продолжительность жизни населения.

Исходной информацией для проведения исследований служили данные о социально-экономических показателях деятельности субъектов РФ, представленные на сайте и в Статистических сборниках Росстата РФ [1]. Используются данные за 7-летний период с 2010 по 2016 годы. Причем в связи с тем, что в Статистическом сборнике (2017) в информации за 2016 год отсутствовали сведения о 2-х (из 135) показателях «Валовой региональный

продукт» и «Валовой региональный продукт на душу населения», то все предварительные расчеты выполнялись по данным за 2015 год с последующей перепроверкой на данных за предыдущие годы и на данных за 2016 год, если не требовались сведения о значениях двух упомянутых показателей.

1. Оценка статистической значимости определяющих факторов

В процессе содержательного анализа социально-экономических показателей, характеризующих деятельность субъектов РФ, из общей совокупности выделено несколько десятков и проведена оценка наличия корреляционной связи с показателем «Ожидаемая продолжительность жизни при рождении, число лет:». Проверка *наличия корреляции* осуществлялась по Р. А. Фишеру на основании t -распределения с $(n-2)$ степенями свободы при $n=85$. Количество показателей с отклоненной нулевой гипотезой оказалось более трех десятков. Коэффициенты корреляции положительные и отрицательные.

На следующем этапе нами исследованы возможности построения функциональной зависимости между показателем «Ожидаемая продолжительность жизни при рождении, число лет» (функция отклика) и статистически с ним связанными показателями (показатели со статистически значимыми коэффициентами корреляции использованы в качестве независимых переменных). Построение регрессионных уравнений при проведении экспериментов, направленных на поиск *лучшей* по статистическим критериям модели, осуществлялось с включением в состав входных переменных различных сочетаний факторов-показателей. Правда, сделать это, как оказалось, совсем не просто: ведь даже при 20-ти элементах число сочетаний по 2, 3, ... элемента весьма велико.

Однако большинство построенных регрессионных моделей при наличии статистически значимых b -коэффициентов недостаточно хорошо аппроксимировало исходные статистические данные ($R^2 < 0,5$). Например, регрессионные модели, включающие различные сочетания показателей «Численность студентов, обучающихся по программам бакалавриата, специалитета, магистратуры, тыс. человек», «Число зарегистрированных преступлений на 100000 человек населения:», «Число больничных коек на 10 000 человек населения», «Численность среднего медицинского персонала на 10 000 человек населения», «Численность врачей на 10 000 человек населения» и др., включая использование объединенного показателя «Образование» (*Численность обучающихся по образовательным программам начального, основного и среднего общего образования + Численность студентов, обучающихся по программам подготовки квалифицированных рабочих, служащих*)* k_1 +(*Численность студентов, обучающихся по программам подготовки специалистов среднего звена*)* k_2 +(*Численность студентов, обучающихся по программам бакалавриата, специалитета, магистратуры*)* k_3 ; $k_1+k_2+k_3=1$), даже при статистически значимых b -коэффициентах имели низкие коэффициенты R^2 и F -критерия.

Поэтому для построения прогнозных моделей использован универсальный алгоритм ранжирования объектов [2–3], позволяющий оперативно выявлять взаимосвязи между объектами, формировать подгруппы с примерно одинаковыми величинами оценочных показателей и анализировать причины образования таких подгрупп, расширять для системы верхнего уровня возможности оптимального выбора определенного объекта, предоставляя сведения о том, *какие объекты и по каким показателям* занимают лидирующее положение, а каждому из объектов обеспечить возможность *анализа резервов* улучшения деятельности, дать первоначальную оценку его конкурентным позициям, рассчитывать статистические характеристики (математическое ожидание, дисперсию, коэффициент вариации, медиану, асимметрию, эксцесс), таблицы и гистограммы распределения, например, административно-территориальных образований (АТО) в зависимости от численных значений показателей,

оценивать динамику доли успешно функционирующих объектов, давать объективную оценку результатам деятельности системы верхнего уровня.

Последующий анализ и построение регрессионных моделей выполнялось нами при использовании исходной информации, представленной в *относительных* единицах.

В окончательном варианте нами получены *статистически значимые* регрессионные модели, *хорошо описывающие* исходную информацию, с составом факторов, включение которых в уравнение регрессии легко обосновать содержательно.

Модель без свободного члена имеет вид:

$$Y = b_1 * X_1 + b_2 * X_2 + b_3 * X_3,$$

где:

Y — относительное значение ожидаемой продолжительности жизни населения административно–территориального образования (по отношению к среднему значению показателя у всей совокупности субъектов РФ);

X_1 — относительный уровень образования в АТО (суммарное относительное значение количества учащихся);

X_2 — относительное количество медицинских работников (просуммированы относительные значения показателей «Численность врачей на 10 000 человек населения» и «Численность среднего медицинского персонала на 10 000 человек населения»);

X_3 — относительное значение показателя «Число больничных коек на 10 000 человек населения».

В результате расчетов получены следующие численные значения статистических показателей качества моделей:

1) для моделей с тремя факторами — независимыми переменными, (по годам):

*значения факторов представлены в относительных единицах

2015 г.

Значения b -коэффициентов:

$b_1=1,115$; $b_2=23,72$; $b_3=17,53$;

$b_1/\sigma_{b_1}=3,64$; $b_2/\sigma_{b_2}=5,88$; $b_3/\sigma_{b_3}=2,27$;

$R^2=0,97$; $R^2_{\text{скадр}}=0,967$; $F_{\text{кр}}=1323,1$.

2016 г.

$b_1=0,0158$; $b_2=0,325$; $b_3=0,27$;

$b_1/\sigma_{b_1}=3,785$; $b_2/\sigma_{b_2}=5,89$; $b_3/\sigma_{b_3}=2,52$;

$R^2=0,98$; $R^2_{\text{скадр}}=0,968$; $F_{\text{кр}}=1375,8$.

2014 г.

$b_1=0,0138$; $b_2=0,36$; $b_3=0,205$;

$b_1/\sigma_{b_1}=3,056$; $b_2/\sigma_{b_2}=6,03$; $b_3/\sigma_{b_3}=1,8$;

$R^2=0,978$; $R^2_{\text{скадр}}=0,966$; $F_{\text{кр}}=1258,1$.

*значения факторов представлены (измерены) в абсолютной шкале:

$b_1=0,0197$; $b_2=0,31$; $b_3=0,18$;

$b_1/\sigma_{b_1}=4,77$; $b_2/\sigma_{b_2}=6,38$; $b_3/\sigma_{b_3}=1,8$;

$R^2=0,98$; $R^2_{\text{скадр}}=0,97$; $F_{\text{кр}}=1487,5$.

2) для моделей с двумя факторами — независимыми переменными:

2016 г.

*значения факторов представлены в относительных единицах

X_1 — относительный уровень образования, т. е. относительное количество учащихся *разделено* на относительную численность населения;

X_2 — сумма относительного количества врачей и относительного количества среднего медицинского персонала на 10000 человек населения АТО.

$$b_1=0,103; b_2=0,283;$$

$$b_1 / \sigma_{b_1}=5,14; b_2 / \sigma_{b_2}=6,95;$$

$$R^2=0,98; R^2_{\text{скорр}}=0,97; F_{\text{кр}}=2250,4.$$

Численные значения оценок коэффициентов регрессии и статистических критериев регрессионных моделей, построенных по данным за 2014 и 2015 годы, мало отличаются от рассчитанных по данным за 2016 год.

*значения факторов представлены (измерены) в абсолютной шкале:

2016 год

$$b_1=260,45; b_2=0,192;$$

$$b_1 / \sigma_{b_1}=6,93; b_2 / \sigma_{b_2}=5,32;$$

$$R^2=0,98; R^2_{\text{скорр}}=0,977; F_{\text{кр}}=2716.$$

Какую же пользу от построенных моделей может получить *население страны и лица*, принимающие решения (ЛПР) на всех уровнях управления? В чем *конкретно* заключается *общественная* полезность этих моделей?

По нашему мнению, *полезность* представленных *статистически значимых* уравнений регрессии, описывающих с *ничтожно малыми ошибками* достаточно *достоверную*, как оказалось (см. Раздел 3), исходную информацию, заключается в том, что:

во-первых, ЛПР на всех уровнях управления *смогут*, получив, наконец, *количественно обоснованный перечень определяющих факторов* и выполнив *оптимизационные* расчеты, *выбирать оптимальный, требующий минимальных затрат ресурсов вариант вложения средств для повышения продолжительности жизни населения*: то ли *способствовать* повышению уровня профессионализма и количества медицинских работников и/или *содействовать* созданию хорошо оснащенных медицинских центров, то ли вложить средства на *повышение уровня образованности* населения, либо решить задачу по *снижению ресурсоемкости товаров и услуг* [4–5] и, тем самым, *уменьшить цену фиксированного набора потребительских товаров* (чтобы население смогло приобрести *большее* количество *полезных* для здоровья продуктов) и т. д.

во-вторых, общественность, население АТО, *повысив* уровень своей образованности (после выполнения ЛПР ряда действий, из состава ранее перечисленных) *получит* возможность *реально осознать, «почувствовать»* тот факт, *что ожидаемая продолжительность жизни* детей и внуков *каждого гражданина* страны будет (с большой вероятностью) *постоянно возрастать*.

Обратим внимание на то, что аналогично, определив на начальном этапе исходный перечень показателей, предположительно связанных с показателями, характеризующими уровень и качество жизни населения (например, путем использования экспертного или количественного анализа [6]), можно сформировать статистически значимый перечень определяющих факторов, управление которыми обеспечит рост уровня жизни граждан России, как, впрочем, и любой другой страны мира.

2. Содержательное обоснование состава определяющих факторов.

В результате выполненных расчетов по представленной на сайте Росстата РФ исходной информации построены *статистически значимые* модели, связывающие показатель «Ожидаемая продолжительность жизни населения административно–территориальных образований» с показателями: $X1$ — относительный уровень образования (число учащихся) в АТО; $X2$ — относительное количество медицинских работников; $X3$ — относительное количество больничных коек.

Причем включенные в модель (вернее, «попавшие» в модель статистически обоснованно) факторы–показатели не только весьма убедительно подтвердили свою «статистическую значимость», но и их включение в модель легко обосновывается содержательно. Действительно, разве не очевидно, что если на территории АТО велика *доля* тех, кто учится, то и вполне вероятно, что там гораздо быстрее (в среднем) будут поступать к населению сведения о том, как нужно заботиться о здоровье, о пользе здорового образа жизни, о новых достижениях медицинской науки и др. Рассуждая аналогично, можно содержательно обосновать целесообразность «участия» в составе независимых переменных нового показателя, сформированного в виде *отношения* количества учащихся на территории АТО к численности проживающего в нем населения.

Даже в таком непростом случае, связанном с обоснованием присутствия в модели фактора «Количество больничных коек на 10 тыс. человек населения», эту ситуацию не сложно объяснить: ведь если в АТО *больше* (на 10 тыс. человек населения) больничных коек, то в нем и *большой* по численности коллектив медицинских работников, больше возможностей для обмена профессиональными знаниями, более выражена профессиональная конкуренция, больше возможностей для появления неординарных, творчески настроенных профессионалов.

3. Выявление аномальных наблюдений в массиве исходной информации

В процессе проведения исследований нами использован достаточно большой объем исходной статистической информации: количество субъектов РФ — 85, а показателей, характеризующих их деятельность — 135, т. е. всего более 10 тысяч числовых данных только за один год. Вполне вероятно, что среди такого обилия чисел могут оказаться выбросы и случайные ошибки, недостоверные, сомнительные, аномальные наблюдения.

Как распознать и исключить такие наблюдения? Ведь наличие их в массиве исходной информации может *негативно повлиять на результаты расчета, на выводы, на качество принимаемых решений*.

Поэтому нами реализован поиск по методике [7] аномальных наблюдений в массиве данных Росстата, использованных в качестве исходной информации для разработки регрессионных моделей. В соответствии с упомянутой методикой построены матрицы корреляций и уравнения регрессии.

Анализ матриц корреляции, рассчитанных по каждой группе выделенных факторов–показателей (по 3-м годам, столбцы — субъекты РФ), совместно с анализом остатков в абсолютном и нормированном вариантах позволил обнаружить аномальные наблюдения в исходных статистических данных у 2-х субъектов РФ (среднее значение коэффициентов корреляции у этих АТО отличается на порядок от средних значений у других АТО). Однако такая аномалия у 2-х АТО (из 85) практически не отразилась на качестве построенных регрессионных моделей.

Выводы

В результате выполненных исследований *впервые*:

1. Построены *регрессионные модели для прогнозирования ожидаемой продолжительности жизни населения административно–территориальных образований *лучшего качества* (при оценке по стандартным критериям статистической значимости — $R^2_{\text{скорр}} > 0,9$; $F_{\text{кр}} \gg 100$); *с лучшими прогнозными свойствами, подтвержденными при «перекрестных» оценках (на данных, сформированных в разные годы); *с использованием *открытых официальных статистических данных и одновременно с проверкой на наличие аномальных наблюдений* в массиве исходной информации, содержащем десятки тысяч числовых данных.

2. Включены в состав *независимых переменных* построенных регрессионных моделей факторы–показатели, *оригинальные по сформированной структуре; *содержательно обоснованные и статистически значимые (в большинстве случаев у b –коэффициентов отношение $b_i/\sigma_{b_i} \gg 2$).

3. Обеспечена возможность, базируясь на сформированном в процессе исследований подмножестве значимых факторов, *проводить оптимизационные расчеты с целью минимизации затрат ресурсов* на оптимальный выбор и реализацию проектов, направленных на *повышение значений показателя «Ожидаемая продолжительность жизни населения административно–территориальных образований»*

Статья подготовлена по результатам исследований, выполненных при поддержке Российского фонда фундаментальных исследований (РФФИ) — проект 18-010-00806/18 «Уровень жизни населения административно-территориальных образований: выявление, исследование, анализ и оценка значимости определяющих факторов (для последующей оптимизации в условиях ограниченных ресурсов)».

Автор благодарен Даниилу Коротину и Дмитрию Сидоренко за подготовку исходных данных, позволившую существенно уменьшить трудозатраты на выполнение расчетов.

Список литературы:

1. Регионы России. Основные характеристики субъектов Российской Федерации. Статистический сборник. М., 2017. Режим доступа: <https://goo.gl/nSRi5r>.
2. Хубаев Г. Н. Ранжирование объектов по множеству количественных показателей: универсальный алгоритм // РИСК: Ресурсы, информация, снабжение, конкуренция. 2018. №1. С. 213-217.
3. Хубаев Г. Н. Качество жизни населения административно-территориальных образований: методика экспресс-анализа // Системный анализ в проектировании и управлении (САЕС-2018): сборник научных трудов XXII Международной научно-практической конференции (г. Санкт-Петербург, 22-24 мая 2018 г.). Т. 2. СПб.: Изд-во Политехн. ун-та, 2018. С. 139-146.
4. Хубаев Г. Н., Калугян К. Х., Родина О. В., Щербаков С. М., Широбокова С. Н. Универсальное методическое и инструментальное обеспечение экспресс-оценки и оптимизации ресурсоемкости товаров и услуг // Бюллетень науки и практики. 2016. №12. С. 286-299.
5. Хубаев Г. Н., Щербаков С. М., Широбокова С. Н. Алгоритмы и программное обеспечение экспресс-оценки ресурсоемкости товаров и услуг // Содружество (Научный российско-китайский журнал). 2016. №9. С. 160-167.
6. Khubaev G. Expert review: method of intuitively agreed choice // 5th International Conference “Economy modernization: new challenges and innovative practice” (November 12, 2017, Sheffield, UK). P. 65-80.

7. Хубаев Г. Н. Способ выявления ошибок в больших массивах числовой информации // Вопросы статистики. 2014. №10. С. 20-25.

*Работа поступила
в редакцию 15.08.2018 г.*

*Принята к публикации
20.08.2018 г.*

Cite as (APA):

Khubaev, G. (2018). Regression models for forecasting life period of population of administrative-territorial education: construction and evaluation of quality. *Bulletin of Science and Practice*, 4(9), 206-217.

Ссылка для цитирования:

Хубаев Г. Н. Регрессионные модели для прогнозирования продолжительности жизни населения административно-территориальных образований: построение и оценка качества // Бюллетень науки и практики. 2018. Т. 4. №9. С. 206-217. Режим доступа: <http://www.bulletennauki.com/khubaev-g-n> (дата обращения 15.09.2018).