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

**REGRESSION MODELS FOR FORECASTING LIFE PERIOD OF POPULATION OF ADMINISTRATIVE-TERRITORIAL EDUCATION: CONSTRUCTION AND EVALUATION OF QUALITY**

*Аннотация.* Построены статистически значимые регрессионные модели (*R*2скорр>0,9; Fкр>>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

*Abstract.*Built statistically significant of the regression model (R2> 0.9, Fcr >> 100; the response function – the Expected life expectancy of the population). Using the original method, the presence of anomalous observations in the array of initial statistical information containing tens of thousands of numerical data was checked.

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

**Введение. 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](http://hdr.undp.org/)) и с 1990 года используется ООН в ежегодном отчёте по развитию человеческого потенциала. Индекс измеряет достижения страны с точки зрения продолжительности жизни, получения образования и фактического дохода, по трём 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](http://gtmarket.ru/ratings/rating-countries-gdp/rating-countries-gdp-info)) 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]. Использованы 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)\*k1+(Численность студентов, обучающихся по программам подготовки специалистов среднего звена)\*k2+(Численность студентов, обучающихся по программам бакалавриата, специалитета, магистратуры)\*k3; k1+k2+k3=*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+*b3\*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; *b3*=17,53;**

***b*1/Ϭb1 =3,64; *b*2/Ϭb2=5,88; *b3* /Ϭb3=2,27;**

***R*2=0,97; *R*2скорр=0,967; Fкр=1323,1.**

**2016 г.**

***b*1=0,0158; *b*2=0,325; *b3*= 0,27;**

***b*1/Ϭb1=3,785; *b*2/Ϭb2=5,89; *b3*/Ϭb3=2,52;**

***R*2 =0,98; *R*2скорр=0,968; Fкр=1375,8.**

**2014 г.**

***b*1=0,0138; *b*2=0,36; *b3*=0,205;**

***b*1/Ϭb1=3,056; *b*2/Ϭb2=6,03; *b3* /Ϭb3=1,8;**

***R*2 = 0,978; *R*2скорр = 0,966; Fкр=1258,1.**

**\*значения факторов представлены (измерены) в абсолютной шкале:**

***b*1=0,0197; *b*2=0,31; *b3*=0,18;**

***b*1/Ϭb1=4,77; *b*2/Ϭb2=6,38; *b3* /Ϭb3=1,8;**

***R*2 = 0,98; *R*2скорр = 0,97; Fкр=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/Ϭb1=5,14; *b*2/Ϭb2=6,95;**

***R*2= 0,98; *R*2скорр = 0,97; Fкр=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/Ϭb1=6,93; *b*2/Ϭb2=5,32;**

***R*2= 0,98; *R*2скорр = 0,977; Fкр=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: ***X*1** – относительный уровень образования (количество учащихся) в АТО; ***X*2** – относительное количество медицинских работников; ***X*3**– 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 *больший* 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.** Насколько нам известно, 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*2скорр>0,9; Fкр >>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* построенныхрегрессионных моделей факторы-показатели, \**оригинальные* *по* сформированной структуре; \*содержательно обоснованные и статистически значимые (в большинстве случаев у *b*–коэффициентов отношение ***bi*/Ϭ*bi*** **>>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. 2017: Статистический сборник. – М., 2017. URL: http://www.gks.ru/free\_doc/new\_site/region\_stat/sep\_region.html

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 (SAEC-2018): Сборник научных трудов XXII Международной научно-практической конференции (г. Санкт-Петербург, СПбПУ им. Петра Великого, 22-24 мая 2018 г.). Том 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 ресурсоёмкости 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 ресурсоёмкости 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). р. 65-80.

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

--------------------------------------------

**ОПУБЛИКОВАНО:**

*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: (дата обращения 15.09.2018). DOI:10.5281/zenodo.1418761