АНАЛИТИЧЕСКИЙ ОТЧЁТ

«Анализ факторов ожидаемой продолжительности жизни по регионам  
средствами линейного моделирования»

В статьях были найдены следующие факторы с потенциально высокой объяснительной способностью зависимой переменной «Ожидаемая продолжительности жизни» (далее ОПЖ):

|  |  |  |
| --- | --- | --- |
| **Карелина и др.** | **«…экономический рост…»** | **«…качество жизни…»** |
| ВВП на душу населения; | Уровень (динамика) ВРП | Объем ВРП |
| Потребление алкоголя, табака, сливочного масла, овощей и фруктов; | Индивидуальные факторы региона (образ жизни) |  |
| Степень урбанизации; |  | Уровень освоенности территории по числу городов на 100 тыс. км2; |
| Уровень безработицы; |  | Уровень безработицы; |
| Среднедушевые фармацевтические расходы;  +  Среднедушевые расходы на здравоохранение (исключая лекарства); |  |  |
| Численность врачей на 1000 человек; |  |  |
| Расходы на образование; |  |  |
| Уровень младенческой смертности; |  |  |
|  | Уровень (сокращение) загрязнения почвы, воздуха, воды |  |
|  |  | Объем вкладов (депозитов) физических лиц в банках в расчете на одного жителя; |
|  |  | Среднее время поиска работы; |
|  |  | Плотность железнодорожных путей общего пользования, километры путей на 10 тыс. км2 территории; |
|  |  | Плотность автомобильных дорог общего пользования с твердым покрытием, километры дорог на 1000 км2 территории; |
|  |  | Доля населения с высшим образованием в общей численности населения в возрасте 15 лет и более; |
|  |  | Миграционный прирост (убыль) населения, тыс. человек; |
|  |  | Число больничных коек на 10 тыс. населения; |
|  |  | Удельный вес общей площади, оборудованной водопроводом; |
|  |  | Удельный вес общей площади, оборудованной водоотведением (канализацией); |
| Обменный курс; |  |  |
| Уровень грамотности; |  |  |
| Среднее число лет обучения для людей старше 25 лет; |  |  |
|  |  | Доля прибыльных предприятий; |
|  |  | Оборот малых предприятий в расчете на одного жителя, тыс. руб |
|  |  | Обеспеченность автозаправочными станциями, число станций на 1000 км дорог; |
|  |  | Выпуск из аспирантуры на 10 тыс. населения; |
|  |  | Число студентов, принятых в образовательные учреждения высшего профессионального образования, на 1000 населения; |
|  |  | Мощность амбулаторно-поликлинических учреждений на 10 тыс. населения; |
|  |  | Оборот розничной торговли на одного жителя; |
|  |  | Численность зрителей театров на 1000 населения. |

Переменные из строк таблицы, выделенных зелёным цветом, были использованы в ходе дальнейшей работы; выделенные красным цветом строки не включались в анализ, поскольку полностью или частично повторяли уже рассмотренные факторы либо казались крайне специфичными (следовательно, их использование предполагалось в случае неудовлетворительности моделирования на допущенных переменных).

Для учёта возможного лага во влиянии переменных на ОПЖ для каждого фактора были собраны данные по всем наблюдённым годам до 2000 года. Полученный набор данных включал в себя следующее (в столбцах – факторы, в строках – года, в которые по данному фактору собиралась статистика):

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Валовой региональный продукт на душу населения** | **Уровень урбанизации** | **Розничные продажи алкогольной продукции на душу населения** | **Среднедушевые денежные доходы населения** | **Уровень безработицы** | **Обеспеченность врачами** | **Младенческая смертность** | **Плотность автомобильных дорог общего пользования с твердым покрытием** | **Средняя сумма накоплений** | **Миграционный прирост** | **Удельный вес общей площади жилых помещений, оборудованной одновременно водопроводом, водоотведением, отоплением, горячим водоснабжением, газом или напольными электроплитами** | **Обеспеченность больничными койками** |
| **2000** |  |  |  |  | + |  |  |  |  |  |  |  |
| **2001** |  |  |  |  | + |  |  |  |  |  |  |  |
| **2002** |  |  |  |  | + |  |  |  |  |  |  |  |
| **2003** |  |  |  |  | + |  |  |  |  |  |  |  |
| **2004** |  |  |  |  | + |  |  |  |  |  |  |  |
| **2005** |  |  |  |  | + |  |  |  |  |  |  |  |
| **2006** |  |  |  |  | + |  |  |  |  |  |  |  |
| **2007** |  |  |  |  | + |  |  |  |  |  |  |  |
| **2008** |  |  |  |  | + |  |  |  |  |  |  |  |
| **2009** |  |  |  |  | + |  |  |  |  |  |  |  |
| **2010** |  |  |  |  | + |  |  |  |  |  |  | + |
| **2011** |  |  |  | + | + |  |  |  |  |  |  |  |
| **2012** |  |  |  | + | + |  |  |  |  |  |  |  |
| **2013** |  |  |  | + | + |  |  |  |  |  |  |  |
| **2014** |  |  |  | + | + |  |  |  |  |  |  |  |
| **2015** |  |  |  | + | + |  |  |  |  |  |  |  |
| **2016** | + |  |  | + | + |  |  |  |  |  |  |  |
| **2017** | + |  |  | + | + |  |  |  |  |  |  |  |
| **2018** | + |  |  | + | + |  |  | + |  |  |  |  |
| **2019** | + |  |  | + | + |  |  | + |  |  |  |  |
| **2020** | + |  |  | + | + |  |  | + |  |  | + |  |
| **2021** | + |  |  | + | + |  | + | + |  |  | + |  |
| **2022** | + | + | + | + | + | + | + | + | + | + | + |  |
| **2023** |  | + | + | + | + |  | + | + | + |  | + |  |
| **2024** |  | + |  |  |  |  | + |  |  |  |  |  |

Корреляционный анализ полученных переменных друг с другом и зависимой переменной ОПЖ позволил установить:

факторы (из отобранных), наиболее сильно коррелирующие с ОПЖ:

|  |  |
| --- | --- |
| **Фактор** | **r** |
| Дороги 2023 | 0,59 |
| Алкоголь 2023 | -0,56 |
| Водопровод+Канализация 2022 | 0,55 |
| Обесп. Больн. Койками 2010 | -0,46 |
| Мигр. Прирост 2022 | 0,45 |

при использовании приведённых выше факторов мультиколлинеарности не наблюдается (наиболее высокий abs(r) между факторами – 0,54).

После определения наиболее сильно коррелирующих с ОПЖ факторов массива данных и установления отсутствия мультиколлинеарности было произведено последовательное построение моделей с увеличением числа используемых в них переменных (переменные добавлялись в модель в порядке убывания модуля коэффицента корреляции фактора и ОПЖ). Результат построений:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Набор факторов** | **R^2** | **F** | **Значимость F** | **Значимость b-коэффициентов** | **A** |
| Дороги | 0,35 | 2,2E-09 | Уравнение значимо | Значим, значим | 2,13% |
| Дороги + Алкоголь | 0,50 | 2,0E-13 | Уравнение значимо | Значим, значим, значим | 1,78% |
| Дороги + Алкоголь + Водопровод | 0,58 | 5,6E-16 | Уравнение значимо | Значим, значим, значим, значим | 1,67% |
| Дороги + Алкоголь + Водопровод + Койки | 0,63 | 2,0E-17 | Уравнение значимо | Значим, значим, значим, значим, значим | 1,65% |
| Дороги + Алкоголь + Водопровод + Миграции | 0,6 | 6E-18 | Уравнение значимо | Значим, значим, значим, значим, значим | **1,51%** |
| Дороги + Алкоголь + Водопровод + Койки + Миграция | **0,65** | 8,4E-18 | Уравнение значимо | Значим, значим, значим, значим, не значим, значим | 1,55% |

Данные таблицы показывают, что наибольшей объяснительной мощностью обладает пятифакторная модель (R2=0,65), однако коэффициент уравнения для включённой в неё переменной «Обеспеченность больничными койками 2010» не значим; исключение данного фактора из уравнения линейной регрессии позволило снизить среднюю ошибку модели (с 1,55% до 1,51%), однако снизилась и доля объяснённой дисперсии (с 0,65 до 0,6). Интерпретировать потерю объясняющей мощности переменной «Обеспеченность больничными койками 2010» можно следующим образом: высокие значения данного фактора, с одной стороны, показывают, что для заболевших людей имеются места для лечения, следовательно, существует вероятность восстановления их здоровья, поэтому ОПЖ растёт; с другой стороны, низкие значения «Обеспеченности» свидетельствуют о высоком уровне здоровья населения в целом, следовательно, в расходные пункты бюджета расширение число больничных коек не включают, в то же время ОПЖ из-за общего здоровья населения высокая (об этом говорят и 95% значения «Обеспеченности больничными койками 2010» при построении пятифакторной модели: от повышения ОПЖ на 0,0016 при увеличении «Обеспеченности» на 10000 чел. на единицу до понижения на 0,039). по совокупности параметров было принято решение оставить полученную после исключения рассмотренного фактора модель как рабочую.

**Вид рабочей модели:**

**ОПЖ = 71,59 + 0,00128 \* Дороги 2023 - 0,472 \* Алкоголь 2023 + 0,058 \* ВодопроводИКанализация 2022 + 0,146 \* Мигр. Прирост 2022**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Коэффициенты** | **Стандартная ошибка** | **P-Значение** | **Нижние 95%** | **Верхние 95%** |
| Y-пересечение | 71,59348 | 1,080681 | 1,17E-71 | 69,44286 | 73,7441 |
| Дороги 2023 | 0,001281 | 0,000507 | 0,013529 | 0,000271 | 0,00229 |
| Алкоголь 2023 | -0,47189 | 0,075808 | 2,09E-08 | -0,62275 | -0,32102 |
| Водопровод+Канализация 2022 | 0,058173 | 0,013742 | 6,1E-05 | 0,030825 | 0,085521 |
| Мигр. Прирост 2022 | 0,146313 | 0,03893 | 0,000324 | 0,06884 | 0,223786 |

Проведены тесты на выполнение моделью условий Гаусса-Маркова (следовательно, на соответствие модели Best Linear Unbiased Estimation). Выполнены три условия из четырёх; значения критерия Дарбина-Уотсона показывают, что автокорреляция остатков, получаемых в ходе использования модели на датасете, не определена, но и не отвергнута, следовательно, о выполнении или невыполнении четвёртого условия сказать нельзя.

**Интерпретация коэффициентов модели:**

* Y-пересечение: с вероятностью 95% региональный ОПЖ при нулевых значениях всех остальных факторов будет лежать в диапазоне от 69,44 лет до 73,74 лет;
* Дороги 2023: с вероятностью 95% увеличение плотности автомобильных дорог общего пользования с твердым покрытием на 1 километр на 1000 км2 территории региона в предыдущем году приведёт к росту регионального ОПЖ в текущем году на 0,0003 - 0,0023 лет (крайне малые значения, однако они поднимают R2 модели (были проведены сравнения моделей с «Дорогами» и без «Дорог», первые всегда оказывались лучше), следовательно, «Дороги» могут быть использованы как фактор, вносящий финальные корректировки в модель); прямая связь плотности дорожного покрытия на ОПЖ может быть объяснена следующим образом: наличие дорожного покрытия, покрывающего большую площадь, позволяет людям, остро нуждающимся в медицинской помощи, быстрее прибыть в больничный пункт (или вообще прибыть туда), следовательно, их вероятность выздороветь растёт;
* Алкоголь 2023: с вероятностью 95% увеличение розничных продаж алкогольной продукции на душу населения на 1 литр в предыдущем году приведёт к падению регионального ОПЖ в текущем году на 0,32 - 0,623 года (от 4 до 7,5 месяцев); данный фактор может быть интерпретирован так: повышение потребления алкоголя приводит к падению ОПЖ в силу вреда алкогольных напитков для здоровья;
* Водопровод+Канализация 2022: с вероятностью 95% увеличение удельного веса общей площади жилых помещений, оборудованной одновременно водопроводом, водоотведением, отоплением, горячим водоснабжением, газом или напольными электроплитами двумя годами ранее приведёт к росту регионального ОПЖ в текущем году на 0,031 - 0,086 лет (на 11 – 31 дней); прямая связь данной переменной и ОПЖ может быть объяснена через улучшение санитарных условий посредством обеспечения населения водопроводом, канализацией и т.п., что позитивно сказывается на здоровье;
* Мигр. Прирост 2022: с вероятностью 95% увеличение миграционного прироста на 1 человека на 1000 человек постоянного населения региона двумя годами ранее приводит к повышению регионального ОПЖ в текущем году на 0,069 - 0,224 лет (на 1 – 3 месяца); такое соотношение может быть объяснено следующим образом: к миграции (если она не вынужденная) способен человек с нормально-хорошим состоянием здоровья, следовательно, приезд здоровых мигрантов повышает общий уровень здоровья населения региона.