

УДК 51

ПОСТРОЕНИЕ РЕГРЕССИОННОЙ ЗАВИСИМОСТИ ОЖИДАЕМОЙ ПРОДОЛЖИТЕЛЬНОСТИ ЖИЗНИ ОТ РЯДА СТАТИСТИЧЕСКИХ ФАКТОРОВ

Протопопова К. В., студентка гр. БЭс-151

Грибанов Е.Н., к.т.н., доцент

Кузбасский государственный технический университет
г. Кемерово

Средняя ожидаемая продолжительность жизни - один из важнейших показателей уровня социально-экономического развития государства, отражающий состояние окружающей среды в стране, здоровье ее населения, бытовые и прочие условия жизни, качество самой жизни населения. По уровню продолжительности жизни по сравнению с экономически развитыми странами мира Россия далеко не лидер, а обострение демографической ситуации, связанное с падением уровня рождаемости, повышением уровня смертности и естественной убылью населения на рубеже XX и XXI веков, еще более усугубило это положение.

За последние годы средняя ожидаемая продолжительность жизни населения России только уменьшалась, что следует признать тревожным признаком. Разрыв между показателями, к примеру, 1987, и 2011 года составляет около 3,5 лет. Для 1994 – 1995 гг. ее падение превышало 5 лет. В настоящее время подобный негативный процесс не характерен для экономически развитых стран и даже среди развивающихся государств является редким случаем.

Средняя ожидаемая продолжительность жизни населения зависит от большого количества различных факторов социально-экономического, биологического и экологического характера. Для России, страны с высокой территориальной дифференциацией климатических и социально-экономических условий жизни населения, ее детальный статистический анализ вызывает большой практический интерес.

Исследуется набор из семи объясняющих результативный признак факторов с числом наблюдений $n = 144$. Число наблюдений выбрано с учетом рассмотрения по регионам, всего приведено 72 регионов Российской Федерации. Данные скомпонованы из данных за два года, 2008-2009, все показатели за год.

Для анализа вида зависимости между результативным признаком y и каждым из семи факторов были проанализированы шесть типов зависимостей (в т.ч. линейная), коэффициенты для которых определялись методом наименьших квадратов. Критерий выбора зависимости – наименьшая остаточная дисперсия.

В результате линейная зависимость между фактором и результативным признаком выбрана оптимальной в последнем случае (x_7). Логарифмическая

и степенная виды зависимостей выбраны как самые удачные для большей части факторов, также выбрана гиперболическая и экспоненциальная зависимость (для факторов x_1 и x_2 соответственно). Нужно отметить, что наименьшей дисперсией в данном анализе обладает степенная зависимость между фактором x_4 и результативным признаком. В данном случае расчетные значения признака наиболее близки к фактическим, следовательно, можно предположить, что фактор x_4 имеет наибольшее влияние на результативный признак.

Высокая корреляция между факторами делает невозможным определение их изолированного влияния на результативный признак. Для выявления мультиколлинеарности использовалась матрица парных коэффициентов корреляции. При анализе мультиколлинеарности были удалены два фактора x_1 (x_1 и x_3).

Для составления уравнения регрессии использовался метод наименьших квадратов, полученная система линейных уравнений была решена по правилу Крамера.

Присутствие аномальных значений, существенно снижающих качество уравнения, обусловлено сильным влиянием в отдельных случаях ряда неизвестных частных факторов. Для выявления аномальных значений было применено правило трех сигм, где в роли математического ожидания выступило выборочное среднее от квадратов разности между фактическими и расчетными значениями результативного признака. После удаления выявленных пяти аномальных значений, уравнение множественной регрессии приняло вид:

$$y = -51,57 + 1,4882 \cdot \ln x_2 + 62,4456 \cdot x_4^{-0,0634} - 0,4510 \cdot \ln x_5 + \\ 63,7875 \cdot x_6^{-0,0058} - 0,0307 \cdot x_7$$

где x_2 – число браков за год, ед.; x_4 – число абортов на 100 родов, ед.; x_5 – число экологических преступлений, ед.; x_6 – число лесных пожаров, ед.; x_7 – обеспеченность населения врачами на 10тыс. населения, чел.

Оценка модели на тесноту совместного влияния объясняющих признаков и на значимость в целом показала, что сила связи между факторами и результативным признаком высока ($R^2 0,7395 > 0,7$) и уравнение в целом значимо (по критерию Фишера $F = 72,6577$, что во много раз превышает критическое значение равное 3,1635). Значимость отдельных факторов проверялась также по критерию Фишера, в результате чего было определено, что все факторы значимы так как расчетные значения ($F_{x_2} = 101,555$; $F_{x_4} = 87,9629$; $F_{x_5} = 22,3961$; $F_{x_6} = 27,9825$; $F_{x_7} = 7,7134$) превышают критическое $F_{kp} = 6,8371$.

Исходя из значения множественного коэффициента детерминации, можно сделать вывод, что 74% изменений результативного признака проис-

ходит по причине изменения рассматриваемых факторов, в остальных 26% случаев на изменение результативного признака влияют другие, не описанные в данной модели, факторы.

На основе приведенных расчетов можно сделать вывод о наибольшем влиянии на результативный признак фактора количества браков. Положительный коэффициент в уравнении регрессии свидетельствует о том, что с увеличением этого фактора результативный признак растет. Зависимость логарифмическая, следовательно, жизнь в браке положительно сказывается на ее продолжительности.

Также достаточно значим фактор количества абортов на 100 родов. Коэффициент в уравнении регрессии имеет положительный знак, однако, зависимость между фактором и признаком обратная, следовательно, с увеличением количества абортов продолжительность жизни снижается. Как и предполагалось в анализе при выборе вида зависимости, фактор абортов оказывает довольно сильное влияние на объясняемый признак.

Фактор экологических преступлений также негативно влияет на результативный признак, знак перед коэффициентом в уравнении отрицательный. Таким образом, совершая экологические преступления, мы укорачиваем свою жизнь. Однако, этот фактор менее значим в построенной модели, и можно предположить, что на самом деле ухудшение экологической обстановки не столь сильно влияет на общую продолжительность жизни, как, например факторы количества браков и абортов. Фактор лесных пожаров, чуть более значим, чем экологический фактор, также оказывает отрицательное влияние на продолжительность жизни, несмотря на выдвинутое предположение об обратном. По-видимому, выделение продуктов горения в атмосферу только ухудшает экологическую обстановку и никаким косвенным образом не может продлить человеку жизнь. Пожары нужно предотвращать, не иначе.

Казалось бы, с увеличением числа врачей, наша жизнь станет долгой. Однако фактор обеспеченности врачами оказался практически на границе по признанию его значимым. Средняя продолжительность жизни человека в современном мире не имеет более столь тесной связи с достижениями в медицине, как в двадцатом веке. Кроме того, коэффициент в уравнении регрессии отрицателен, следовательно, количество врачей не всегда служит знаком высокого качества медицинской помощи.

Было получено уравнение множественной регрессии линейное по структуре с нелинейными компонентами. Уравнение значимо в целом, а также значимы все факторы. Теснота связи между факторами и результативным признаком высока, что свидетельствует о хорошем качестве уравнения и адекватности построенной модели. Все факторы вошли в полученное уравнение со знаками, не противоречащими действительности и логическим предположениям о направленности влияния.

Список литературы:

- 1) Кремер, Н.Ш. Эконометрика: учебник для вузов./ Н. Ш. Кремер, Б.А. Путко. – М.: ЮНИТИ-ДАНА, 2002. – 311 с.
- 2) Вучков И., Бояджиева Л., Солаков Е., Прикладной регрессионный анализ. – М.: Финансы и статистика, 1987, 239 с.
- 3) Иберла К. Факторный анализ.— М.: Статистика, 1980.—308 с.
- 4) Гаврилов Л.А.. Гаврилова Н.С. Биология продолжительности жизни. М.: Наука, 1991, 280 с.
- 5) Продолжительность жизни: анализ и моделирование - сборник статей под ред. Андреева Е.М., Вишневского А.П. - М: Статистика, 1979 г. - 157 с.