

Математические методы анализа миграционных процессов на основе демографических данных

Бутусов О.Б., Никифорова О.П., Редикульцева Н.И.
 Московский государственный машиностроительный университет (МАМИ),
 Государственный университет управления,
 АНО ВПО "Московский гуманитарный университет",
 (495)682-2053, butusov-1@mail.ru, (499)177-5413, olga.nikiforova2013@yandex.ru,
 (499) 374-6010, redik_ni@mail.ru

Аннотация. Рассмотрена задача математического анализа демографических данных, в которых информация о миграционных потоках учитывается неявным образом. Для математического анализа демографических процессов использованы методы регрессионного, нейросетевого и стохастического анализов. Рассмотрены две возрастные группы населения: молодые (0 – 39 лет) и пожилые (40 – 70 лет). При разработке стохастической модели использована теория цепей Маркова и матрица переходных вероятностей. Параметризация и идентификация модели проведены по данным Росстата.

Ключевые слова: математическая модель, стохастическое моделирование, демографические процессы, миграционные потоки, Марковские цепи, переходные вероятности, метод наименьших квадратов

Проблема анализа и прогнозирования демографических процессов имеет важное социологическое значение. В настоящее время разработано достаточно большое количество математических демографических моделей [1 – 3], имеющих большой прогнозный потенциал.

Разделяя население на подгруппы, например согласно возрасту, и учитывая взаимодействие подгрупп, можно получить модели механизмов, описывающих динамические процессы переходов между возрастными группами. Данные модели могут быть полезны для получения оценки прироста населения или его снижения.

Для прогнозирования демографических процессов нами исследованы три группы моделей: регрессионные, нейросетевые и стохастические.

Регрессионная модель позволяет построить математическую функцию, например, в виде полинома, с помощью которой удобно прогнозировать тренды временных рядов. На рисунке 1 представлена динамика подгруппы молодые для населения России и ее регрессионная модель, полученная с помощью полинома четвертого порядка. При этом абсолютная среднеквадратическая ошибка модели составляет $\Delta y = 0.00033$.

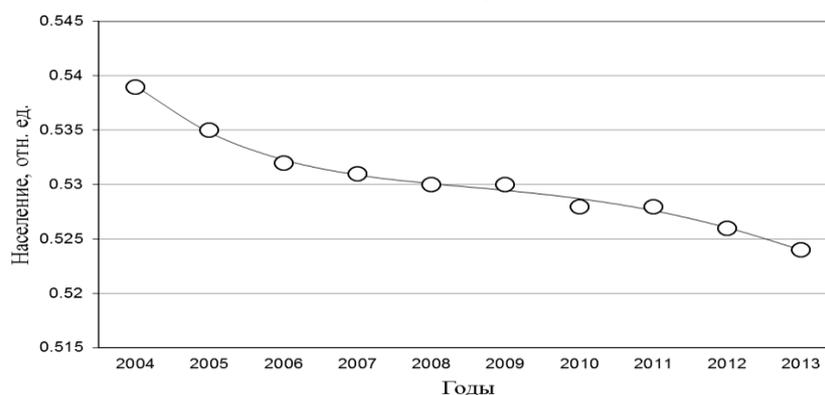


Рисунок 1. Динамика подгруппы "молодые" населения России и ее регрессионная модель в виде полинома четвертого порядка

Для различных степеней регрессионного полинома получаются различные регрессионные модели. При этом график величины абсолютной среднеквадратической ошибки имеет

вид, представленный на рисунке 2. Из рисунка следует, что насыщение достигается при степени полинома равной четырем. После этого увеличение степени полинома не приводит к улучшению результатов моделирования.

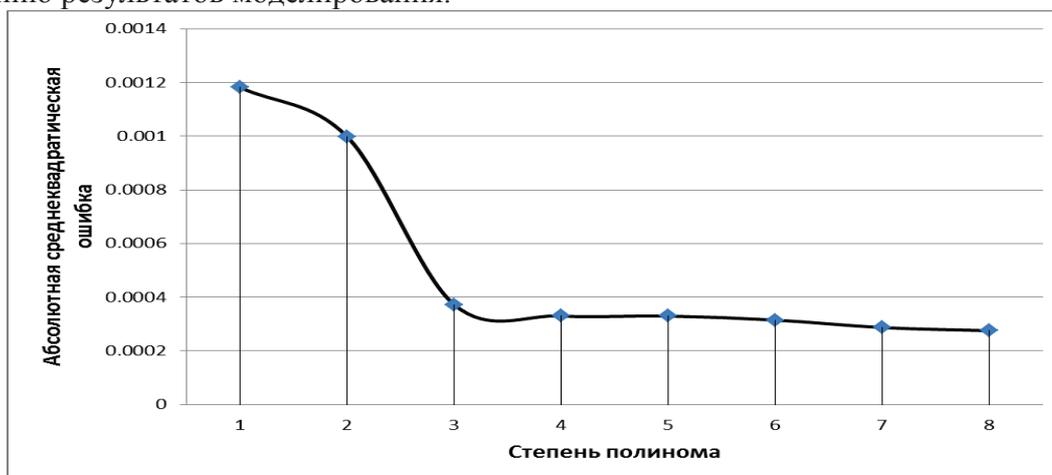


Рисунок 2. График абсолютной среднеквадратической ошибки регрессионных полиномиальных моделей в зависимости от степени полинома

Известно, что для прогнозирования временных рядов широко используются нейросетевые модели и нейронные сети [4]. Рассмотрим результат моделирования статистических данных, представленных на рисунке 1, с помощью простейшей нейросетевой модели. Для моделирования используем единственный линейный нейрон. Схема модели представлена на рисунке 3.

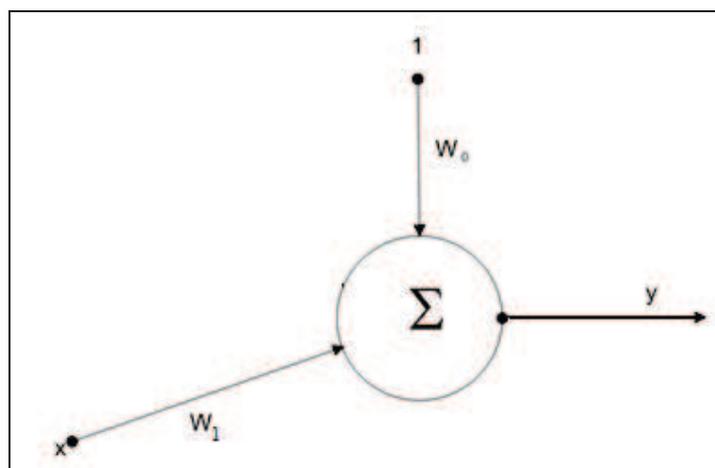


Рисунок 3. Нейросетевая модель, использованная для моделирования динамики подгруппы "молодые": x – годы, w_0, w_1 – весовые коэффициенты, y – относительная численность подгруппы "молодые"

Для тестирования модели данные были разделены на две подгруппы: обучающая подгруппа размером восемь значений и контрольная размером два значения. В результате был получен прогноз контрольных значений с ошибками: 0.0066072 и 0.0092131, а также прогноз на следующий 2014-й год для относительной численности подгруппы "молодые": $M = 0.53382$.

Также была разработана стохастическая модель, которая идентифицирована по данным Росстата для периода 2004 – 2013 годы.

Для построения стохастической модели была использована теория цепей Маркова, основу которой составляют переходные вероятности. Переходной вероятностью p_{ij} называют условную вероятность того, что из состояния i в результате воздействия различных факторов, таких как смертность, рождаемость, миграция и пр. система переходит в состояние j . Та-

ким образом, в обозначении p_{ij} первый индекс указывает номер предшествующего, а второй – номер последующего состояний. Например, p_{23} – вероятность перехода из второго состояния в третье.

Положим, что число состояний системы конечно и равно k . Матрицей перехода системы называют матрицу, которая содержит все переходные вероятности этой системы:

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1k} \\ p_{21} & p_{22} & \cdots & p_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ p_{k1} & p_{k2} & \cdots & p_{kk} \end{pmatrix}. \quad (1)$$

Так как в каждой строке матрицы помещены вероятности событий (перехода из одного и того же состояния i в любое возможное состояние j), которые образуют полную группу, то сумма вероятностей этих событий равна единице. Другими словами, сумма переходных вероятностей каждой строки матрицы перехода равна единице.

Для расчета переходных вероятностей был использован метод наименьших квадратов (МНК).

Рассмотрим следующую задачу. Разделим население на две группы: молодые (0 – 39 лет) и пожилые (40 – 70 лет). Исследуемый интервал 2004 – 2013 гг. разделим на восемь временных интервалов. Численность молодого и пожилого населения на каждом из временных интервалов k будем оценивать с помощью вероятностей P_{1k} и P_{2k} , а неизвестные переходные вероятности рассчитаем по статистическим данным с помощью МНК. При этом матрица перехода между временными интервалами будет иметь следующий вид:

$$P = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}. \quad (2)$$

Введем более простые обозначения для переходных вероятностей:

$$\begin{pmatrix} P_1 \\ P_2 \end{pmatrix}_k = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} \cdot \begin{pmatrix} P_1 \\ P_2 \end{pmatrix}_{k-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} P_1 \\ P_2 \end{pmatrix}_{k-1}. \quad (3)$$

При этом функционал МНК примет следующий вид:

$$F(a, b, c, d) = \sum_{k=1}^8 (P_{1k} - aP_{1,k-1} - bP_{2,k-1} + P_{2k} - cP_{1,k-1} - dP_{2,k-1})^2 \Rightarrow \min. \quad (4)$$

Задача оптимизации (4) приводит к следующей системе уравнений:

$$\frac{\partial F}{\partial a} = 0, \quad \frac{\partial F}{\partial b} = 0, \quad \frac{\partial F}{\partial c} = 0, \quad \frac{\partial F}{\partial d} = 0. \quad (5)$$

Система уравнений (5) факторизуется на следующие две системы одинаковой структуры:

$$\begin{aligned} a \sum_k P_{1,k-1}^2 + b \sum_k P_{1,k-1} P_{2,k-1} &= \sum_k P_{1k} P_{1,k-1}, \\ a \sum_k P_{1,k-1} P_{2,k-1} + b \sum_k P_{2,k-1}^2 &= \sum_k P_{1k} P_{2,k-1}, \\ c \sum_k P_{1,k-1}^2 + d \sum_k P_{1,k-1} P_{2,k-1} &= \sum_k P_{2k} P_{1,k-1}, \\ c \sum_k P_{1,k-1} P_{2,k-1} + d \sum_k P_{2,k-1}^2 &= \sum_k P_{2k} P_{2,k-1}. \end{aligned} \quad (6)$$

Аналитическое решение систем уравнений (6) имеет следующий вид:

$$\begin{aligned}
 a &= \frac{\sum_k P_{1k} P_{1,k-1} \cdot \sum_k P_{2k}^2 - \sum_k P_{1k} P_{2,k-1} \cdot \sum_k P_{1,k-1} P_{2,k-1}}{\sum_k P_{1k}^2 \cdot \sum_k P_{2k}^2 - \left(\sum_k P_{1,k-1} P_{2,k-1}\right)^2}, \\
 b &= \frac{\sum_k P_{1k} P_{2,k-1} \cdot \sum_k P_{1,k-1}^2 - \sum_k P_{1k} P_{1,k-1} \cdot \sum_k P_{1,k-1} P_{2,k-1}}{\sum_k P_{1k}^2 \cdot \sum_k P_{2k}^2 - \left(\sum_k P_{1,k-1} P_{2,k-1}\right)^2}, \\
 c &= \frac{\sum_k P_{2k} P_{1,k-1} \cdot \sum_k P_{2k}^2 - \sum_k P_{2k} P_{2,k-1} \cdot \sum_k P_{1,k-1} P_{2,k-1}}{\sum_k P_{1k}^2 \cdot \sum_k P_{2k}^2 - \left(\sum_k P_{1,k-1} P_{2,k-1}\right)^2}, \\
 d &= \frac{\sum_k P_{2k} P_{2,k-1} \cdot \sum_k P_{1,k-1}^2 - \sum_k P_{2k} P_{1,k-1} \cdot \sum_k P_{1,k-1} P_{2,k-1}}{\sum_k P_{1k}^2 \cdot \sum_k P_{2k}^2 - \left(\sum_k P_{1,k-1} P_{2,k-1}\right)^2}.
 \end{aligned} \tag{7}$$

В результате численного решения по формулам (7) была рассчитана матрица переходных вероятностей:

$$P = \begin{pmatrix} 0,534 & 0,463 \\ 0,529 & 0,468 \end{pmatrix}. \tag{8}$$

Поясним некоторую особенность полученной матрицы переходных вероятностей. Из (8) следует, что переходная вероятность из группы пожилые в группу молодые отлична от нуля. Этот эффект обусловлен влиянием миграционных потоков и свидетельствует о том, что миграционные потоки оказывают существенное влияние на современную демографию.

С помощью полученных переходных вероятностей был рассчитан прогноз по численности населения на 2014 год:

$$P^T = \begin{pmatrix} 0,534 & 0,529 \\ 0,463 & 0,468 \end{pmatrix} \begin{pmatrix} 0,534 & 0,529 \\ 0,463 & 0,468 \end{pmatrix} \cdot \begin{pmatrix} 0,52 \\ 0,479 \end{pmatrix} = \begin{pmatrix} 0,518 \\ 0,481 \end{pmatrix}. \tag{9}$$

В результате была получена следующая численная оценка: численность молодого населения РФ в 2014 году должна составить 74253,75 (тыс. человек), а пожилого населения - 68949,91 (тыс. человек).

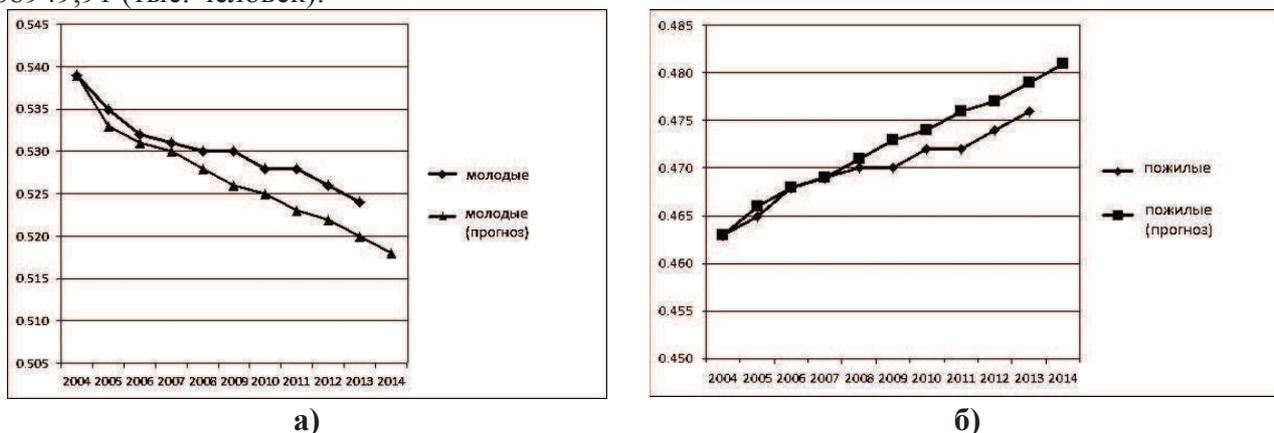


Рисунок 4. Результаты моделирования и данные Росстата по численности населения РФ: а – группа молодого населения, б – группа пожилого населения (погрешность вычисления составила 0,2%)

С помощью модели были рассчитаны теоретические графики динамики численности населения РФ, которые приведены на рисунке 4 вместе с данными Росстата.

Выводы

1. Сравнение результатов моделирования с исходными данными демонстрирует достаточно хорошее соответствие модельных расчетов и статистических данных для всех трех рассмотренных моделей.
2. Миграционные потоки учитываются в моделях косвенным образом, однако полученные результаты свидетельствуют о хороших суммарных предсказательных способностях модели.
3. Найденная матрица переходных вероятностей стохастической модели учитывает миграционные потоки и может быть предложена для прогнозирования демографических процессов.

Литература

1. Пыров П.В., Бутусов О.Б. Математическое и компьютерное моделирование миграционных процессов / Труды института системного анализа РАН (ИСА РАН). Динамика неоднородных систем. М.: Книжный дом "Либроком", 2010. т. 50(1). – С. 219 – 222.
2. Павловский Ю.Н. Моделирование, декомпозиция и оптимизация сложных динамических процессов. М.: Издательство ВЦ РАН, 2009. – 139 с.
3. Белотелов Н.В., Бродский Ю.И., Кручина Е.Б. и др. Имитационная игра на основе эколого-демографо-экономической модели (ЭДЭМ): описание и инструкция пользователю. М.: РХТУ им.Д.И. Менделеева, 2003. – 83 с.
4. Саймон Хайкин. Нейронные сети. – М.: Вильямс, 2006. – 1104 с.