

УДК 519.25, 314.143

МОДЕЛИРОВАНИЕ ИНТЕГРАЛЬНЫХ ПОКАЗАТЕЛЕЙ ДЕМОГРАФИЧЕСКОГО РАЗВИТИЯ НА ПРИМЕРЕ МУНИЦИПАЛЬНЫХ ОБРАЗОВАНИЙ НОВГОРОДСКОЙ ОБЛАСТИ

Жгун Т.В.

ГОУ ВПО «Новгородский государственный университет имени Ярослава Мудрого»,
Великий Новгород, e-mail: zhtv@mail.ru

В статье предложена методика построения интегральной характеристики динамической системы с помощью метода главных компонент на основе измеряемых данных. Показано, что при изменении входных данных все главные факторы определяются разными переменными и применение метода главных компонент при вычислении интегральной характеристики не позволяет произвести корректное сравнение показателей объектов. Рассматривая статистические данные с точки зрения теории обработки сигналов как сигнал, в котором имеется полезная информация – слабый сигнал, который нужно выделить, и шум, и предполагая, что изменение входных данных может иметь как случайную природу (шум), так и неслучайную (сигнал), будем считать, что именно неслучайные коэффициенты главных компонент и являются той характеристикой системы, которая позволит нам определить веса интегрального индикатора и, значит, качество системы на промежутке наблюдений при изменении входных данных. Наличием неслучайного (т.е. значимого) вклада переменной в структуру главных компонент будем считать не большую величину факторной нагрузки, а инвариантность факторной нагрузки при возмущениях, признаком которой будет величина отношения сигнал/шум. Приводится апробация методики на примере вычисления интегральной оценки демографического развития на основе данных государственной статистики для муниципальных образований Новгородской области. Предлагаемая методика может быть использована для вычисления интегральных оценок качества любой динамической системы.

Ключевые слова: региональная демографическая политика, демографические характеристики, мониторинг системы показателей, метод главных компонент, качество системы управления, система показателей эффективности

SIMULATION OF INTEGRAL RATES OF DEMOGRAPHIC DEVELOPMENT BY THE EXAMPLE OF MUNICIPAL DISTRICTS OF NOVGOROD OBLAST

Zhgun T.V.

Novgorod State University. Yaroslav the Wise, Veliky Novgorod, e-mail: zhtv@mail.ru

Methods of building the integral characteristic of dynamic system with principal component analysis based on measured data is proposed in the article. It is shown that with changes in input data all the main factors are defined by different variables and use of principal component analysis does not allow correct comparison of objects' figures while computing integral characteristics. Considering statistical data from the point of view of signal processing as signal consisting of useful information (signal that needs to be extracted), and noise. We assume that change in input data can have either stochastic (noise) or non-stochastic nature (signal). We consider non-stochastic coefficients of principal components to be the characteristic that allows us to determine the weights of integral indicator and therefore the quality of system on observation period with changes in input data. We consider non-stochastic (i.e. significant) contribution to the structure of principal components to be made not by bigger factor load, but invariance of factor load during perturbation, the indication of which is value of signal-to-noise ratio. Approximation of methods by the example of computing the integral estimation of demographic development of municipal districts of Novgorod Oblast by the state statistical data is presented. Proposed methods can be used for computing integral estimations of quality of any dynamic system.

Keywords: regional demographic policy, demographic characteristics, monitoring system performance, the method of principal components, quality management system, system of parameters of efficiency

Проблемы демографического развития страны в последнее время стали позиционироваться как наиболее существенные для поступательного развития России. Демографические процессы, которые недавно были особо значимыми лишь для специалистов, в настоящее время выдвинулись в центр общественного внимания. Государством в последние годы предпринимаются меры, направленные на изменение негативных демографических процессов.

Для реализации государственной демографической политики необходимо своевременное реагирование на демографические тенденции в текущий период. Обеспечение

своевременного реагирования немислимо без учета региональных и территориальных особенностей демографического развития, для которого необходима организация системы мониторинга демографической ситуации в Российской Федерации и ее субъектах, а также создание необходимых инструментов такого мониторинга. Объективное определение характеристик демографического развития является обязательным инструментом такого мониторинга. Целью работы является создание инструмента мониторинга демографической ситуации в Российской Федерации – методика объективного определения характеристик демографического развития.

Традиционно демографическая ситуация характеризуется по трем основным параметрам: рождаемости, смертности и миграции. Но при всей важности и необходимости использования отдельных показателей, они не могут обеспечить адекватное сравнение разных территорий. Трудности сравнения между собой регионов по характеру демографической ситуации не только в ограниченных возможностях использования для этого тех или иных показателей. Часто противоположная направленность динамики различных показателей в разных регионах не позволяет однозначно оценивать ситуацию. Так, в Москве в 2005 г. суммарный коэффициент рождаемости составлял 1,140, а ожидаемая продолжительность жизни – 71,36 года, тогда как в Эвенкийском автономном округе соответственно 2,261 и 57,56. Очевидно, что когда уровни и динамика демографических показателей различны, невозможно произвести сравнение территорий [5]. Следовательно, для выбора основных точек управленческого воздействия необходим анализ демографической динамики в целом.

По мнению С.А. Айвазяна, интегральный индикатор той или иной синтетической категории качества представляет собой «свертку» оценок более частных свойств и критериев этого понятия: статистических показателей либо экспертных оценок [1]. Отметим, что применение экспертных оценок наряду с субъективностью имеет ещё один недостаток: не для всех оцениваемых систем можно найти достаточное количество квалифицированных экспертов. В частности, такова и рассматриваемая проблема: экспертов, одинаково хорошо разбирающихся в тонкостях демографической ситуации в различных муниципальных образований, видимо, не существует вообще. Следовательно, использование статистических данных остаётся единственно возможным. Необходимо отметить, что при формировании структуры интегрального показателя и его расчете возникают серьезные затруднения, связанные с субъективностью выбора первичной информации, с достоверностью и репрезентативностью статистической информации и с методической неопределенностью определения значения интегрального показателя [1, 3, 4]. Отметим, что переменные, значения которых можно измерять только статистически, имеют для исследуемой ситуации достаточно условный характер, лишь опосредованно отражая внутреннюю структуру рассматриваемой интегральной характеристики.

Рассмотрим построение интегральной оценки системы из m объектов, для

которой в моменты времени $t = t_1, t = t_2, \dots, t = t_k$ известны таблицы описаний этих объектов размерностью $m \times n$ – матрицы $A^k = \{a_{ij}^k\}_{i,j=1}^{n,m}$, $k = 1, 2, \dots, p$. Для каждого момента $t = t_k$ вектор интегральных показателей будет иметь вид

$$q^k = A^k \cdot w^k, \quad (1)$$

где $q^k = \langle q_1^k, q_2^k, \dots, q_m^k \rangle^T$ – вектор интегральных индикаторов; $w^k = \langle w_1^k, w_2^k, \dots, w_m^k \rangle^T$ – вектор весов показателей для момента. Для фиксированного момента $t = t_k$ интегральную оценку чаще всего записывают для каждого рассматриваемого объекта с номером i в виде аддитивной свертки данных с некоторыми весами

$$q_i = \sum_{j=1}^n w_j \cdot a_{ij}, \quad i = 1, 2, \dots, m. \quad (2)$$

Обычно веса w_j назначаются экспертами. Метод экспертных оценок широко используется в силу простоты получения информации, но не может считаться объективным. Лишены субъективности формальные методы, в частности, широко используется метод главных компонент, предложенный С.А. Айвазяном [1]. Согласно [1], вектор интегральных индикаторов для каждого объекта $q = A \cdot w_1$ есть проекция векторов-строк матрицы данных A на первую главную компоненту, w_1 – собственный вектор, соответствующий максимальному собственному значению ковариационной матрицы $\Sigma = A^T \cdot A$.

Этот метод используется для оценки статических систем в случае, если первая главная компонента хорошо приближает моделируемую ситуацию, т.е. в случае, если максимальное собственное число ковариационной матрицы дает вклад не менее 70% в сумму всех собственных чисел. Однако если такое предположение не выполняется, проекцию на первую компоненту нельзя считать удачной оценкой. Ситуацию можно исправить, если вместо одной компоненты выбрать l компонент, чтобы относительная доля разброса γ_l , приходящаяся на первые l ($l \leq n$) l (главных компонент):

$$\gamma_l = \frac{\lambda_1 + \lambda_2 + \dots + \lambda_l}{\lambda_1 + \lambda_2 + \dots + \lambda_n} \geq \theta, \quad (3)$$

была не менее определенной величины. Мы будем рассматривать далее случай, когда $\theta = 0,75$. Выбрав главные компоненты, в случае, когда результаты применения метода главных компонент (или метода факторного анализа) устраивают исследо-

вателя, для каждого признака эффект воздействия выбранных факторов суммируем и таким образом определяем веса в (2). Критерием удовлетворительного решения обычно является возможность четкой содержательной интерпретации получившихся факторов. Метод главных компонент формально не требует использования вращений, так как предлагает единственное решение; но в практике обычно используется вращение варимакс для облегчения интерпретации компонент, который приводит к увеличению больших и уменьшению малых значений факторных нагрузок.

Рассмотрим статистические данные, предоставленные Комитетом государственной статистики Великого Новгорода, отражающие демографическую ситуацию в муниципальных образованиях Новгородской области. Среди имеющихся данных была устранена мультиколлинеарность и рассматривались 9 переменных: коэффициент рождаемости; коэффициент смертности; коэффициент младенческой смертности; коэффициент брачности; коэффициент раз-

водимости; коэффициент миграции; доля населения моложе трудоспособного возраста; доля населения старше трудоспособного возраста; доля мужчин среди населения. В табл. 1 приведены результаты факторного анализа данных, относящихся к различным моментам наблюдения.

Интегральная оценка вычисляется на основании нормированных данных (т.е. все значения приведены на интервал $[0, 1]$ и соответствуют принципу «чем больше, тем лучше»: большему значению показателя (при прочих равных) соответствует большее значение интегрального индикатора). Предполагается, что в данных нет объектов-выбросов и пропущенных значений [2]. Для вычисления главных компонент выполнено центрирование, после которого данные имеют нулевое среднее и единичное среднеквадратичное отклонение. Данные были обработаны с помощью пакета *SPSS Statistics* версии 11.0 и для каждого года построены главные факторы. Крестиками отмечены значимые переменные – те, у которых абсолютное значение факторной нагрузки не менее 0,6.

Таблица 1

Результаты факторного анализа

| 1 фактор | | | | | | | | | |
|------------------|---|---|---|---|---|---|---|---|---|
| Номер переменной | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 2006 | | | | + | + | | | | + |
| 2008 | | + | | + | + | | | + | |
| 2009 | | + | | | + | | | | |
| 2010 | | + | | | | | | + | |
| 2 фактор | | | | | | | | | |
| 2006 | | + | | | | | | + | |
| 2008 | + | | | | | | + | | |
| 2009 | + | | | + | | | | + | |
| 2010 | | | | + | + | | | | |
| 3 фактор | | | | | | | | | |
| 2006 | | + | + | | | | | | |
| 2008 | | | | | | | | | |
| 2009 | | | | | | | | | |
| 2010 | | | + | | | | | | |
| 4 фактор | | | | | | | | | |
| 2006 | + | | | | | | | | |
| 2008 | | | | | | | | | |
| 2009 | | | | | | | + | | |
| 2010 | + | | | | | | | | |

Значимые переменные для всех факторов разных лет сильно отличаются. Значит, невозможно верно выделить значимые (определяющие ситуацию) переменные для этих факторов для всей совокупности данных, в то время как анализ данных любого года в отдельности позволяет это сделать. Чтобы иметь возможность оценивать характеристики системы в динамике, посмотрим на получаемую при работе информацию с точки зрения теории обработки сигналов

как на сигнал, в котором имеется полезная информация – слабый сигнал, который нужно выделить, и шум. Измерение переменных неизбежно связано с точностью измерительного прибора, поэтому любой полученный с помощью измерения результат неизбежно содержит ошибку измеряемых данных, которая носит случайный характер. Статистические данные – тоже результат измерения, неизбежно содержащего погрешность измерения. Любой результат, по-

лученный на основании этих данных, также будет содержать эту неустранимую ошибку. Переход к другому моменту времени означает изменение данных, которое вызвано как изменением объясняющих переменных, так и изменением случайной составляющей. Весовые коэффициенты при построении интегрального показателя будут иметь совсем иные значения для следующего момента наблюдений. Причем изменение этих значений может быть вызвано как изменением входных данных, так и случайными ошибками. Предполагая непрерывную зависимость вычисляемого показателя от изменения входных данных, определим природу такого изменения.

Будем считать, что именно истинные коэффициенты главных компонент и являются той характеристикой системы, которая позволит нам определить веса интегрального индикатора и, значит, качество системы на промежутке наблюдений. Наличием неслучайного (т.е. значимого) вклада переменной в структуру главных компонент будем считать не большую величину факторной нагрузки, а инвариантность факторной нагрузки при возмущениях, признаком которой будет величина отношения сигнал/шум. По статистическим данным для ряда наблюдений определим веса факторных

нагрузок для l первых (наиболее весомых) главных компонент и определим значение интегрального показателя для фиксированного момента времени $t = t_k$ по формуле

$$q^k = A^k \cdot W_l, \quad (4)$$

где $W_l = \langle w_1, w_2, \dots, w_l \rangle$ – матрица, составленная по первым l компонентам с использованием всех наблюдений с учетом (3).

Применим изложенную методику для вычисления интегральной оценки демографического развития муниципальных образований Новгородской области. В табл. 2 представлено определение неслучайных коэффициентов факторных нагрузок для второго фактора. В последней строке отмечены те переменные, где факторную нагрузку можно считать неслучайной – переменные, где отношение сигнал/шум более чем 2,2. Такая величина применяется в технике при выделении сигнала из сильно зашумленных данных. Этому условию отвечают переменные 4, 5 (коэффициенты брачности и рождаемости) и переменные 6 (коэффициент миграции) и 9 (доля мужчин). Направление собственных векторов выбираем, чтобы максимизировать услышанный сигнал, т.е. сумму ОСШ у действующих переменных.

Таблица 2

Коэффициенты факторных нагрузок для второго фактора

| Фактор 2 | Факторные нагрузки | | | | | | | | | |
|----------|--------------------|-------|-------|-------|-------------|--------------|--------------|-------|-------|--------------|
| Годы | Собств. числа | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 2000 | 1,65 | 0,24 | -0,20 | -0,27 | 0,22 | -0,19 | 0,66 | -0,08 | -0,01 | -0,55 |
| 2006 | 1,45 | -0,13 | -0,21 | -0,17 | 0,35 | -0,29 | -0,25 | -0,22 | -0,32 | -0,70 |
| 2008 | 1,97 | -0,61 | 0,17 | -0,34 | 0,12 | -0,26 | -0,18 | -0,56 | -0,01 | -0,22 |
| 2009 | 1,76 | -0,47 | -0,02 | 0,37 | 0,14 | -0,26 | -0,38 | -0,48 | -0,17 | -0,38 |
| 2010 | 1,54 | 0,19 | -0,32 | -0,41 | 0,21 | -0,46 | -0,33 | -0,03 | -0,11 | -0,55 |
| среднее | 1,67 | -0,16 | -0,12 | -0,16 | 0,21 | -0,29 | -0,10 | -0,27 | -0,12 | -0,48 |
| разброс | 0,20 | 0,37 | 0,18 | 0,33 | 0,09 | 0,11 | 0,07 | 0,25 | 0,11 | 0,19 |
| ОСШ | 8,21 | 0,45 | 0,62 | 0,58 | 2,37 | 3,29 | 4,03 | 1,05 | 1,28 | 2,58 |

Значения весовых коэффициентов переменных, полученные по 4 главным компонентам с учетом выбора направлений компонент, максимизирующего целевую функцию с и соответствующей нормировки на корень из собственных чисел, согласно (3), приведены в табл. 3.

Демографические данные – это, пожалуй, единственный вид статистических данных, для которых значение интегральной оценки ожидаемо. Очевидно, что в любом субъекте России налицо одна и та же ситуация: на фоне некоторого общего улучшения демографической ситуации в рассматриваемый период имеется неравномерность положения в регионе. Центр

является «пылесосом», выкачивающим из глубинки молодое активное население, из-за чего положение в центре самое лучшее, а дальние районы представляют «кольцо смерти», окружающее благополучный центр. Ближайшие к центру районы и районы с развитой промышленностью занимают промежуточное положение между благополучным центром и неблагополучными окраинами. Именно такую картину и дают нам вычисленные значения интегрального показателя. Изменение значения интегрального показателя по годам для некоторых районов Новгородской области можно увидеть на рисунке. Близкие к центру районы (Старорусский) и районы с развитой

промышленностью (Парфинский) занимают промежуточное положение между Великим Новгородом и отдаленными районами (Батецкий район).

Таблица 3

Суммарные значения весовых коэффициентов

| Рожд. | Смертн. | Млад. смертн. | Брачн. | Развод-ть | Миграция | Доля молодежи | Доля старш. возр. | Доля мужчин |
|-------|---------|---------------|--------|-----------|----------|---------------|-------------------|-------------|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 0,872 | 0,823 | 0,477 | 1,077 | -1,059 | 0,000 | 0,000 | 0,895 | -0,210 |

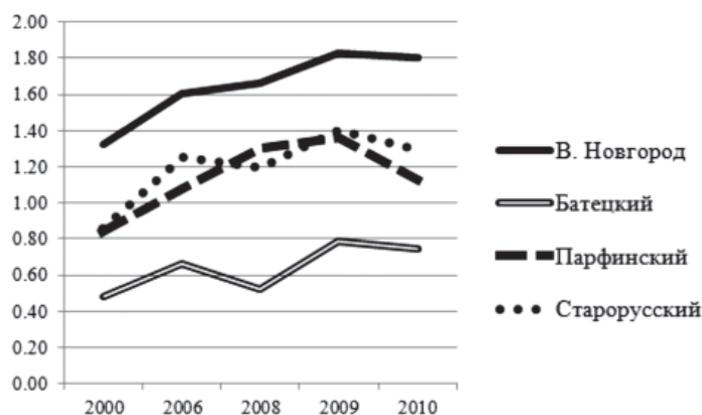


График изменения демографической интегральной характеристики

Характеристики демографической ситуации той или иной территории дают лишь интегральное представление о характере демографических процессов, причем лишь на фоне общей совокупности рассматриваемых территорий. Они показывают, где совокупно демографические показатели хуже, а где лучше относительно некоторого общего среднего, которым может быть и средний индекс по региону, и его медианное значение. Вычисление интегральных оценок демографического развития не исключает необходимости тщательного анализа всех его составных элементов и поиска путей их улучшения.

Заключение

Предлагаемая методика может быть использована при организации реализации государственной демографической политики в территориальных образованиях Российской Федерации любого уровня, для построения интегральной оценки социально-экономического развития и для вычисления интегральных оценок изменения качества любой системы.

Список литературы

1. Айвазян С. А. Интегральные индикаторы качества жизни населения: их построение и использование в социально-экономическом управлении межрегиональных сопоставлений. – М.: ЦЭМИ РАН, 2000. – 56 с.
2. Загоруйко Н.Г. Прикладные методы анализа данных и знаний. – Новосибирск: Изд-во Института математики, 1999. – 270 с.
3. Кузнецов М.П., Стрижов В.В. Построение интегрального индикатора с использованием ранговой матрицы описаний // Интеллектуализация обработки информации: доклады 9-й международной конференции, 2012. – С. 130–132.

4. Мельников Д.И., Стрижов В.В., Андреева Е.Ю., Эдензартер Г. Выбор опорного множества при построении устойчивых интегральных индикаторов // Математические методы распознавания образов. 14-я Всероссийская конференция ММРО-14: сборник докладов. – МАКС-Пресс, 2009. – С. 159–162.

5. Рыбаковский В.В. Интегральная оценка демографического неблагополучия регионов России. (Методологические вопросы) // Социальная политика и социальное партнерство. – 2008. – № 6 <http://rybakovsky.ru/stati1a6.html>.

References

1. Ajvazjan S.A. Integrated indicators of the quality of life of the population: their construction and use in socio-economic governance interregional comparisons. Moscow: CEMI RAS, 2000. 56 p.
2. Zagorujko N.G. Applied methods of analysis of data and knowledge. - Novosibirsk: Publishing house of Institute of mathematics, 1999. 270 p.
3. Kuznecov M.P., Strizhov V.V. *Intellektualizacija obrabotki informacii. Doklady 9-j mezhduнародной konferencii* [Intellectualization of information processing], Budva, 2012. pp. 130–132.
4. Mel'nikov D.I., Strizhov V.V., Andreeva E.Ju., J eden-zarter G. *Matematicheskie metody raspoznavanija obrazov. 14-ja Vserossijskaja konferencija MMRO-14.* [Mathematical methods of pattern recognition. The 14th All-Russian Conference MMRO-14]. Moscow, MAX-Press, 2009. pp. 159–162.
5. Rybakovskij V.V. *Social policies and social partnerships.* 2008. no. 6. URL: <http://rybakovsky.ru/stati1a6.html> (date of access: 17.05.13).

Рецензенты:

- Пиковский А.А., д.э.н., профессор кафедры прикладной экономики Новгородского государственного университета имени Ярослава Мудрого, г. Великий Новгород;
Кириянов Б.Ф., д.т.н., профессор кафедры прикладной математики и информатики Новгородского государственного университета имени Ярослава Мудрого, г. Великий Новгород.

Работа поступила в редакцию 19.07.2013.