

## РОЛЬ ОБРАЗОВАНИЯ В ПОДГОТОВКЕ СПЕЦИАЛИСТОВ ДЛЯ ЦИФРОВОЙ ЭКОНОМИКИ

В. А. ЛИВИНСКАЯ, канд. физ.-мат. наук, доц.  
Белорусско-Российский университет  
Могилев, Беларусь

### Аннотация

В статье рассмотрены вопросы применения современных информационных технологий в образовательном процессе на примере использования метода главных компонент при построении рейтинга районных центров Могилевской области.

### Ключевые слова:

Информационные технологии, метод главных компонент, собственные значения, показатели уровня жизни населения.

В условиях цифровизации экономики будущим финансистам, экономистам, менеджерам уже в ближайшее время предстоит использовать не только фундаментальные классические знания, но и навыки практической обработки и анализа больших массивов данных. Умение анализировать и реагировать на изменение внешних факторов должны способствовать решению задач повышения стоимости компании путем увеличения инвестиционной привлекательности, грамотного использования информации о конкурентах, совершенствования качества прогнозирования.

Одной из обязательных компетенций будущего специалиста, занимающегося анализом экономической информации в наше время, является способность использовать готовые программные продукты. Они, как правило, ориентированы на пользователя, не владеющего глубоко алгоритмами машинного обучения и языками программирования, но хорошо представляющего ситуацию, когда их можно использовать. Это под силу современным студентам, которые без гаджетов уже не представляют свою жизнь, и проведение лабораторных занятий с использованием современных программных продуктов будет для них интересной формой обучения.

Изучение дисциплин, связанных с прикладным эконометрическим анализом, должно обязательно сопровождаться возможностью овладеть современными инструментами анализа данных, к которым относится программа «STATISTICA» (производитель StatSoft, USA). Преимуществами этого инструмента по сравнению с наиболее доступным АНАЛИЗОМ ДАННЫХ из EXCEL является то, что в ней содержится полный набор модулей прикладного статистического анализа, качественный визуальный анализ как исходной совокупности, так и результатов.

Начинать знакомство с современными технологиями обработки данных необходимо в университете, осваивая курсы «Информационные тех-

нологии», «Статистика», «Эконометрика», работая с реальными данными. Сбор данных в современных исследованиях занимает до 80 % времени работы над проектом. Конечно, для сбора больших объемов данных (более тысячи наблюдений), необходимо использование специальных инструментов, таких как, например, парсинг сайтов (если информация собирается в открытых источниках ИНТЕРНЕТ), но для отработки практических навыков можно использовать готовую статистическую информацию, опубликованную в открытом доступе.

Для иллюстрации применения использования программы «Statistica» при построении интегральной рейтинговой оценки уровня жизни в 19 районных центрах Могилевской области использовался метод главных компонент. Это один из методов прикладного статистического анализа, решающий задачу снижения признакового пространства и классификацию объектов. Целью исследования являлся анализ динамики рейтингового показателя в 2015 и 2016 гг. Тема «Уровень жизни населения» входит в тематику изучения дисциплины «Статистика» на 2-м курсе.

Исходными данными явилась информация, размещенная в сборнике «Статистический ежегодник Могилевской области – 2017» [1]. В качестве факторов, характеризующих уровень жизни в районных центрах, рассматривались такие показатели, как:

- $X_1$  – товарооборот на душу населения (млн р.);
- $X_2$  – ввод в эксплуатацию жилых домов (тыс. м<sup>2</sup>);
- $X_3$  – номинальная зарплата (р.);
- $X_4$  – занятое население в городах и районах (на конец года; в процентах к численности);
- $X_5$  – количество врачей на 1000 чел.

Поскольку каждый район описывается пятью показателями с различными единицами измерения, для ранжирования районных центров необходимо использовать методы многомерного статистического анализа. Так, для снижения размерности признакового пространства (а именно эта цель и стоит в данном случае) необходимо получить линейную комбинацию исходных признаков, т. е. перейти от матрицы признаков к вектору, являющемуся линейной комбинацией из главных компонент, образующих новую систему координат. Из пяти базисных векторов будут отобраны те, в направлении которых исходные признаки имеют наибольшую дисперсию. Этот алгоритм реализован в модуле «Principal Components and Classification» программы «Statistica».

Первый шаг по переходу к новой системе координат предполагает построение ковариационной матрицы как меры обобщенной вариации всех признаков и вычисление ее собственных значений. Нормированные собственные векторы этой матрицы и выступают в качестве базисных в новой системе координат. Доказано, что сумма собственных значений равна размерности исходного пространства.



Далее отбираются главные компоненты – базисные вектора, собственные значения которых в сумме дают более 70–80 % общей дисперсии. Как правило, они соответствуют собственным значениям, больше единицы. На рисунке 1 представлен график под названием «каменистая осыпь», предназначенный для отображения последовательности собственных значений и отбора на основании этого главных компонент для признаков 2015 г.

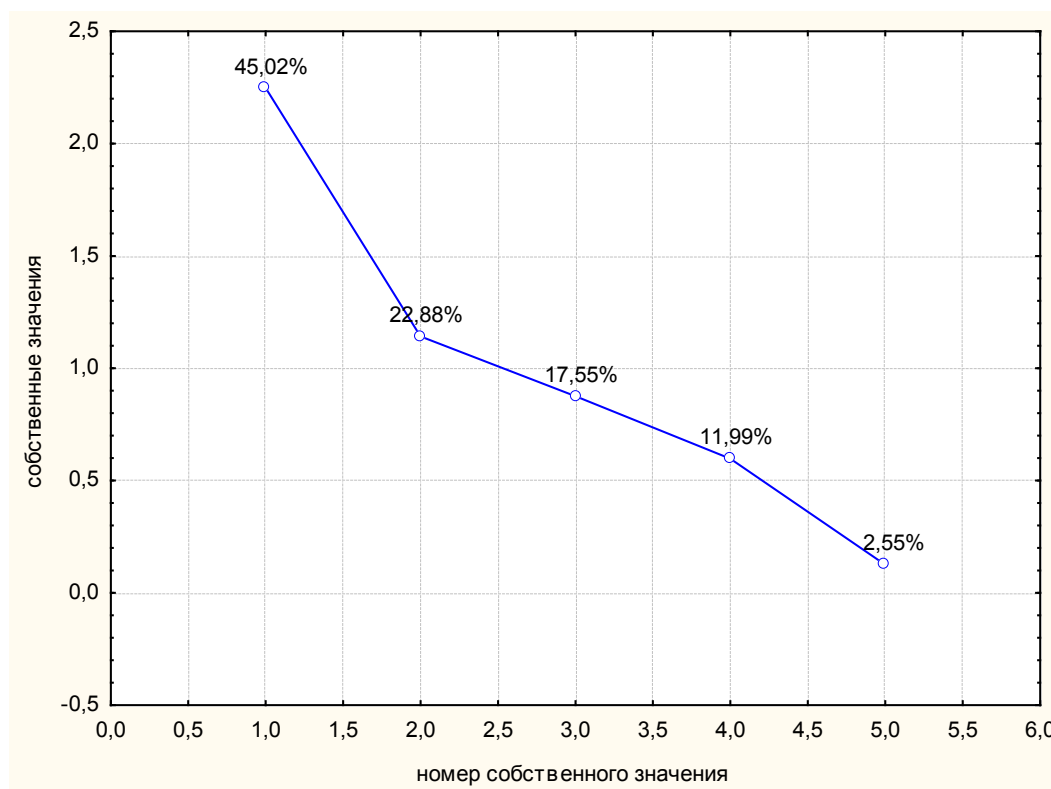


Рис. 1. Собственные значения корреляционной матрицы показателей в 2015 г.

Для определения корреляционной связи между исходными признаками и новыми искусственно построенными главными компонентами программа вычисляет коэффициенты корреляции между ними. Визуально индикатором того, на сколько хорошо каждая из исходных пяти показателей воспроизводится выделенными главными компонентами, качество передачи может служить единичный круг, представленный на рис. 2. Координаты точек являются коэффициентами корреляции. Чем ближе точки к окружности, тем лучше признак воспроизведен в системе координат выбранных факторов.

Анализируя рис. 2, можно отметить, что в первую главную компоненту (Factor1) вошли все исходные показатели, кроме товарооборота на душу населения, причем ввод в эксплуатацию жилых домов ( $X_2$ ) и процент занятого населения ( $X_4$ ) имеют отрицательную связь с ней, а номинальная заработная плата ( $X_3$ ) и количество врачей на 1000 человек ( $X_5$ ) положительно связаны с первой главной компонентой.

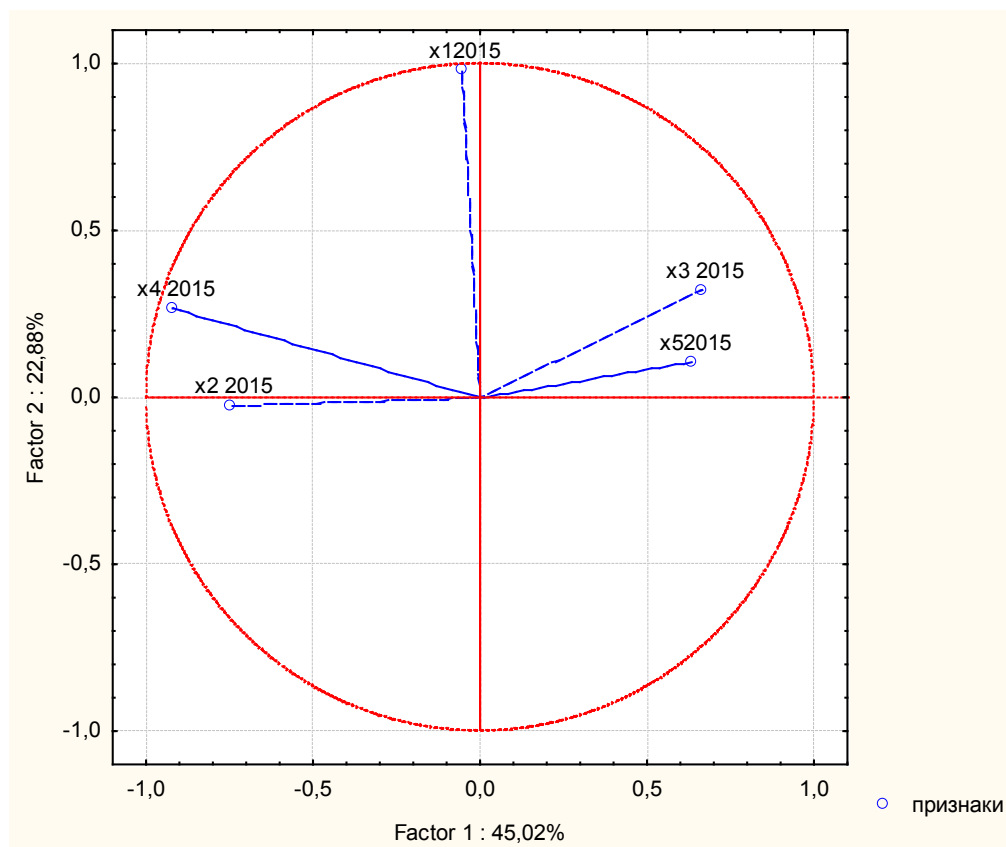


Рис. 2. Проекция признаков на факторную плоскость

Для построения рейтинга районов Могилевской области были найдены координаты каждого районного центра в новой системе координат (состоящей из первой и второй главной компоненты, а затем проведено ранжирование упорядочение по возрастанию). Программа «Statistica» позволяет это сделать несколькими кликами мыши. Так результат ранжирования в 2015 и 2016 гг. по первой главной компоненте представлен на рис. 3.

Результат данного анализа может быть полезен областным властям, в компетенции которых находится мониторинг уровня жизни населения области и отслеживание лидеров и аутсайдеров. Так, в 2016 г. по сравнению с 2015 г. очень сильно поднялся в рейтинге Бельничский район (с 10 на 4 место), в этот же период Хотимский район опустился с 12 на 18 место. Подробный анализ причин, вызывающих такую динамику, позволит воздействовать на ситуацию.

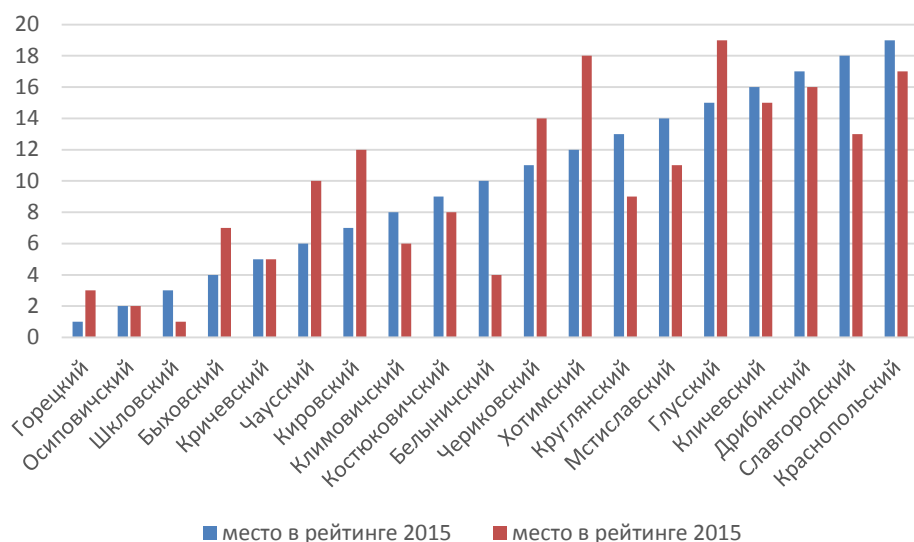


Рис. 3. Динамика рейтинга районных центров по первой главной компоненте

Данная методика может быть применена к различным объектам, представляющим многомерную совокупность в ситуациях, когда нужно сжать информацию и использовать ее для принятия управленческих решений. Это могут быть решения по упорядочению объектов как для анализа изменения положения в общем списке (ранжирование по первой главной компоненте), так и для разбиения совокупности на кластеры (т. е. группы объектов, схожие с объектами в своем кластере и различающиеся с объектами других кластеров). Например, разбиение на кластеры заемщиков банка в зависимости от информации, указанной при получении кредита с целью предложения им различных банковских продуктов. Динамическая модификация метода главных компонент (так называемый метод «Гусеница») применяется при прогнозировании поведения финансовых рынков, но ее использование требует более фундаментальной подготовки в области математических наук.

Получение навыков работы с готовыми программными продуктами и реальными данными позволит выпускникам экономических специальностей повысить заинтересованность работодателя на собеседовании при поиске первого места работы, стимулирует их к научно-исследовательской работе, определит тренд в профессиональном росте.

#### СПИСОК ЛИТЕРАТУРЫ

1. Статистика // Национальный статистический комитет Республики Беларусь [Официальный сайт]. – Режим доступа: <http://www.belstat.gov.by/>. – Дата доступа: 16.11.2018.