

Научная статья

Original article

УДК 330.43

doi: 10.55186/2413046X_2023_8_4_183

**СИНТЕЗ АЛГОРИТМА ВЫБОРА РЕАЛИЗАЦИИ МЕТОДОВ
ФАКТОРНОГО АНАЛИЗА ЭКОНОМЕТРИЧЕСКИХ ДАННЫХ В R И
PYTHON**

**SYNTHESIS OF AN ALGORITHM FOR CHOOSING THE
IMPLEMENTATION OF METHODS OF FACTOR ANALYSIS OF
ECONOMETRIC DATA IN R AND PYTHON**



Астафьев Рустам Уралович, ассистент кафедры Высшей Математики и Программирования, Федеральное государственное бюджетное образовательное учреждение высшего образования «МИРЭА – Российский технологический университет»

Пронина Елена Владиславовна, к.ф.-м.н, доцент, доцент кафедры Высшей Математики и Программирования, Федеральное государственное бюджетное образовательное учреждение высшего образования «МИРЭА – Российский технологический университет»

Пихтилькова Ольга Александровна, к.ф.-м.н, доцент, доцент кафедры Высшей Математики и Программирования, Федеральное государственное бюджетное образовательное учреждение высшего образования «МИРЭА – Российский технологический университет»

Параскевопуло Ольга Ригасовна, к.ф.-м.н., доцент кафедры Высшей Математики и Программирования, Федеральное государственное бюджетное образовательное учреждение высшего образования «МИРЭА – Российский технологический университет»

Морозова Татьяна Анатольевна, старший преподаватель кафедры Высшей Математики и Программирования, Федеральное государственное бюджетное образовательное учреждение высшего образования «МИРЭА – Российский технологический университет»

Кузнецова Екатерина Юрьевна, старший преподаватель кафедры Высшей Математики и Программирования, Федеральное государственное бюджетное образовательное учреждение высшего образования «МИРЭА – Российский технологический университет»

Немировская-Дутчак Ольга Эрнестовна, старший преподаватель кафедры Высшей Математики и Программирования, Федеральное государственное бюджетное образовательное учреждение высшего образования «МИРЭА – Российский технологический университет»

Astafiev Rustam Uralovich

Pronina Elena Vladislavovna

Pikhtilkova Olga Alexandrovna

Paraskevopulo Olga Rigasovna

Morozova Tatyana Anatolyevna

Kuznetsova Ekaterina Yurievna

Nemirovskaya-Dutchak Olga Ernestovna

Аннотация. В данной статье приведены результаты проведения синтеза алгоритма выбора реализации методов факторного анализа.

Таким образом, целью работы является синтез алгоритма выбора языка реализации факторного анализа статистического исследования данных путём рассмотрения вариантов реализации факторного анализа в R и Python на примерах некоторых экономических задач, формирования экономической интерпретации результатов данных реализаций, выделения основных характеристик методов.

В качестве основных методов в работе применяются методы сравнительного анализа реализаций в разных языках и выделения их

характеристик, а также метод синтеза для формирования алгоритма подбора реализации в соответствии с основными их характеристиками.

В результате проведения работы был выполнен теоретический обзор реализации методов факторного анализа в R и в Python, выделены основные необходимые характеристики методов, проведён сравнительный анализ на основе характеристик, по которым далее составлена таблица вопросов с оценками, на основе которых и был сформирован алгоритм выбора языка реализации. Был приведён пример ответов на вопросы и общий алгоритм выбора реализации.

По результатам синтеза, можно также сделать вывод о том, что применимость конкретного метода может быть обоснована только внедрением его в последующие исследования, по итогам которых и возможна доработка метода по отдельным характеристикам.

Abstract. This article presents the results of the synthesis of the algorithm for choosing the implementation of factor analysis methods.

Thus, the aim of the work is to synthesize an algorithm for choosing the language for implementing factor analysis of statistical data research by considering options for implementing factor analysis in R and Python using examples of some economic problems, forming an economic interpretation of the results of these implementations, highlighting the main characteristics of the methods.

The main methods used in the work are methods of comparative analysis of implementations in different languages and highlighting their characteristics, as well as a synthesis method for forming an algorithm for selecting an implementation in accordance with their main characteristics.

As a result of the work, a theoretical review of the implementation of factor analysis methods in R and Python was carried out, the main necessary characteristics of the methods were identified, a comparative analysis was carried out based on the characteristics, according to which a table of questions with

estimates was compiled, on the basis of which an algorithm for choosing the implementation language was formed. An example of answers to questions and a general algorithm for choosing an implementation were given.

Based on the results of the synthesis, it can also be concluded that the applicability of a particular method can only be justified by its implementation in subsequent studies, according to the results of which it is possible to refine the method according to individual characteristics.

Ключевые слова: факторный анализ, алгоритм проведения факторного анализа, сравнительный анализ, методы анализа в R, методы анализа в Python, синтез алгоритма выбора метода

Keywords: factor analysis, algorithm of factor analysis, comparative analysis, analysis methods in R, analysis methods in Python, synthesis of method selection algorithm

Введение

Факторный анализ в экономике - это процедура, с помощью которой большее число переменных сокращается за счёт установления связи между ними и сведению их к меньшему числу величин, каждая из которых называется фактором. Получается, что в фактор входят те переменные, которые сильнее других связаны между собой. Переменные из разных факторов при этом слабо коррелируют друг с другом [1].

Помимо практической значимости (фактически, факторный анализ приводит к сокращению числа переменных), факторный анализ позволяет дать экономическую интерпретацию результатов, так как зачастую отдельные переменные не до конца способны отразить реальный экономический эффект от совокупности взаимосвязанных факторов, которые часто скрыты от глаз и измерения непосредственно исследователем.

Таким образом, факторный анализ - важный инструмент, владение которым не только делает практическое исследование качественнее, но и

позволяет работать с его результатами, интерпретировать их с экономической точки зрения не только по смыслу, но и «на языке чисел».

Если обратиться к сайту eLibrary, то по поиску «Факторный анализ» можно увидеть огромное количество трудов, которые включают его в себя, при этом описанные работы касаются самых разных научных областей, что говорит о его универсальности и широкой применимости.

Благодаря своей распространённости факторный анализ стал не только неотъемлемой частью многих исследований, но и получил свою реализацию в различных программах и языках программирования, которые используются, в частности, для проведения статистического анализа.

В данной работе мы рассмотрим применение и реализацию факторного анализа в R и в Python, установим основные характеристики данных реализаций, что позволит нам сформировать общий подход к выбору конкретной реализации факторного анализа в данных языках.

Таким образом, задачи, которые стоят перед нами – рассмотреть варианты реализации факторного анализа в R и Python на примерах некоторых экономических задач, дать экономическую интерпретацию результатов данных реализаций, выделить основные характеристики методов, на основе характеристик сформировать алгоритм выбора реализации факторного анализа для экономических задач.

Материалы и методы

В качестве основных методов в работе применяются методы сравнительного анализа реализаций в разных языках и выделения их характеристик, а также метод синтеза для формирования алгоритма подбора реализации в соответствии с основными их характеристиками.

В качестве материалов используется литература по факторному анализу в экономике [1], [2] и [3], документация по реализации факторного анализа в R [4] и [5] и Python [6], [7], а также примеры реализации методов на открытых исходных данных [8], [9], [10], [11].

Литературный обзор

Факторный анализ в R.

Наиболее очевидная в работе и практически применяемая функция для факторного анализа в R - это функция `factanal ()` [4].

Функция `factanal ()` представляет собой реализацию факторного анализа максимального правдоподобия на основе ковариационной матрицы и/или матрицы данных.

В общем случае вызов функции со всеми параметрами выглядит следующим образом:

```
factanal (x, factors, data = NULL, covmat = NULL, n.obs = NA, subset,
na.action, start = NULL, scores = c ("none", "regression", "Bartlett"), rotation =
"varimax", control = NULL)
```

Рассмотрим каждый из параметров:

`x` – формула, матрица или любой объект, который может быть преобразован в матрицу данных;

`factors` – число факторов, которые необходимо выделить;

`data` – дополнительные даты (в случае, если `x` – формула, применяемая к набору данных);

`covmat` – ковариационная матрица данных;

`n.obs` – количество наблюдений, если задана ковариационная матрица;

`subset` – подмножество, если необходимо выделить часть данных из исходных;

`na.action` – спецификация действия, если `x` – формула;

`start` – NULL, либо матрица начальных значений;

`scores` – тип генерации весов факторов: Томпсона, либо при учёте Бартлетта и проч.

`rotation` – характер вращения факторов, то есть характер изменения матрицы нагрузок;

`control` – список управляющих переменных;

nstart – количество начальных значений, по умолчанию 1;

trace – выводить ли информацию о трассировке в процессе анализа;

lower – нижняя граница уникальности, по умолчанию 0,005;

opt – список управляющих значений, которые будут переданы в качестве аргументов в control;

rotate – список дополнительных аргументов для функции вращения.

Результат работы функции приведён на Рисунке 1.

```
factanal(diffs, factors = 2)
#>
#> Call:
#> factanal(x = diffs, factors = 2)
#>
#> Uniquenesses:
#> APC BP BRY CVX HES MRO NBL OXY ETP VLO XOM
#> 0.307 0.652 0.997 0.308 0.440 0.358 0.363 0.556 0.902 0.786 0.285
#>
#> Loadings:
#> Factor1 Factor2
#> APC 0.773 0.309
#> BP 0.317 0.497
#> BRY
#> CVX 0.439 0.707
#> HES 0.640 0.389
#> MRO 0.707 0.377
#> NBL 0.749 0.276
#> OXY 0.562 0.358
#> ETP 0.283 0.134
#> VLO 0.303 0.350
#> XOM 0.355 0.767
#>
#> Factor1 Factor2
#> SS loadings 2.98 2.072
#> Proportion Var 0.27 0.188
#> Cumulative Var 0.27 0.459
#>
#> Test of the hypothesis that 2 factors are sufficient.
#> The chi square statistic is 62.9 on 34 degrees of freedom.
#> The p-value is 0.00184
```

Рисунок 1 – Листинг выдачи результатов работы функции

Рассмотрим вызов функции и построчно результаты, которые она выдаёт.

Uniquenesse фактически показывает, какой процент заданной переменной может быть объяснён на текущем количестве факторов (которые мы задаём исходно в виде параметра).

Loadings показывает фактические нагрузки переменных в заданных факторах. Данный пункт позволяет нам сделать выводы о вхождении отдельных переменных в состав полученных факторов.

Следующие пункты - таблица объяснённой дисперсии для полученных факторов, а после указано p-value, которое необходимо для оценки полученного результата анализа.

Модель факторного анализа в общем случае выглядит следующим образом:

$$x = \Lambda f + e \quad (1)$$

где x – матрица-вектор на p элементов, Λ – матрица $p \times k$ нагрузок, f – матрица весов на k элементов, а e – p -элементная матрица ошибок.

Фактически, работа функции основывается на создании модели факторного анализа путём вращения корреляционной матрицы в виде подгонки, которая оптимизирует логарифмическую вероятность при нормальности уникальности коэффициентов [5].

Таким образом, факторный анализ в R обеспечивает подбор оптимальной модели, содержащей и учитывающей веса компонентов элементов по смыслу их вхождения в каждый из факторов в том или ином количестве. В оптимальном случае, следует провести факторный анализ для разного числа факторов для оценки успешности модели и подбора наилучшего количества факторов.

Факторный анализ в Python.

Для факторного анализа в Python возможно применение различных комбинаций функций и методов, но в данной работе будет рассмотрена функция FactorAnalyzer, которая включает в себе все основные характеристики и варианты реализации факторного анализа [5].

FactorAnalyzer в Python позволяет выполнить два вида факторного анализа: исследовательский факторный анализ и подтверждающий факторный анализ, то есть, помимо обычной процедуры выделения

факторов, позволяет разделить новые данные по факторам, которые были выделены в предыдущих итерациях для других данных.

Исследовательский факторный анализ проводится в 2 шага: извлечение факторов и их вращение, то есть производится разделение дисперсии по показателям, а затем выделение и преобразование показателей в некоррелированные факторы, чтобы улучшить общую интерпретируемость.

Как правило, для начала факторного анализа необходимо выбрать количество факторов, которые необходимо выделить в процессе дальнейшего анализа. Критерий Кайзера заключается в том, что значительная доля дисперсии должна быть объяснена с помощью выбранных факторов. Так собственное значение является хорошим критерием для количества факторов, так как его значение больше 1 может стать критерием для выбора конечной функции.

Дополнительно, в Python можно провести тесты для факторизируемости исходных данных с помощью тестов Бартлетта и Кайзера-Мейера-Олкина.

При этом в самой функции доступны разные параметры поворота полученной матрицы нагрузок:

- варимакс;
- промакс;
- облимин;
- облимакс;
- квартимин;
- квартимакс;
- эквамакс;
- geomin_obl;
- geomin_ort.

Данная функция, как мы видели, многокомпонентная, а результат зависит от формата представления функции. Основные характеристики, так

или иначе, могут отражать все итоговые характеристики для факторного анализа.

Фактически, функция основывается на использовании \min_{res} , то есть на минимизации нормальной формы корреляционной матрицы в условиях работы с нагрузками и итоговым представлением повернутой матрицы по факторам [6].

Результаты

Общий алгоритм проведения факторного анализа

По итогам рассмотренных реализаций можно сделать вывод об общем алгоритме проведения факторного анализа для обеих реализаций:

1. В исходных данных выделить зависимую и независимые переменные.
2. Установить факторизируемость данных.
3. Определить количество факторов.
4. Провести факторный анализ исходных данных.
5. Оценить качество факторной модели и сделать выводы о результатах анализа.
6. Дать экономическую интерпретацию полученной модели в контексте поставленной задачи.

Данный алгоритм решения задачи проведения факторного анализа, как может выступить решением конкретных задач, так и пунктами оценки характеристик каждой из реализаций.

Примеры реализации факторного анализа, таким образом, будут приведены в рассмотрении их реализации с точки зрения заданного общего алгоритма, что позволит систематизировать их результаты, а также выделить общие паттерны текущих решений.

Для заданных реализаций подготовим основные элементы характеристик языков, по которым будет оцениваться каждая из них.

Таблица 1 – Основные характеристики для оценки реализаций факторного анализа

Характеристика	Описание
Простота использования	Данная характеристика определяет возможности использования функции и сложность их применения, так как, чем сложнее применять конкретные функции, тем менее дружелюбными являются дополнительные изменения функции, необходимые в контексте каждой конкретной задачи.
Простота внедрения	Данная характеристика важна в контексте работы и применения алгоритма анализа в контексте более крупной задачи или конкретного программного решения, так как чем проще внедрение, тем в более сложных программных продуктах возможно их применение.
Распространённость языка	Данная характеристика важна в условиях уникальности задачи в общем контексте решений одного типа задач. Если задача факторного анализа заключается в единственном её применении, либо на одних и тех же наборах данных, то не будет требоваться дополнительное обучение специалиста, либо найм квалифицированного сотрудника в контексте задачи.
Возможности интерпретации	Данная характеристика позволяет оценить, на сколько проста интерпретация результатов только на основе вывода результатов функции, то есть нужна ли дополнительная квалифицированная консультация по результатам, либо дообучение по заданной теме.
Возможности оценки результатов	Данная характеристика заключается в возможности подтверждения результатов анализа и их статистической и экономической интерпретаций.
Простота предварительного анализа данных	Данная характеристика позволяет оценить, на сколько простым является алгоритм предварительного анализа данных и приведение их к виду, при котором можно достичь высокого качества самого анализа.

Таким образом, данный набор характеристик может помочь в оценке реализации факторного анализа в разных языках, с учётом их специфичных характеристик в контексте конкретной задачи.

Пример проведения факторного анализа в R.

Для использования функции в R дополнительные библиотеки не нужны.

Загрузка и обработка файла до необходимого представления набора данных.

```
food <- read.csv("https://userpage.fu-berlin.de/soga/300/30100_data_sets/food-texture.csv",  
               row.names = "x") ##загружаем открытые данные из интернета  
str(food) ##проверяем корректную загрузку данных, выведя информацию о структуре файла
```

Рисунок 2 – Загрузка файла исходных данных в R

В R предварительный анализ не проводился, так как количество факторов подбиралось в соответствии с p-value.

Выполнение анализа.

```
food.fa ##вызываем метод факторного анализа  
food.fa$uniquenesses ##выводим одну конкретную область результатов  
plot(food.fa$loadings[,1], ##построение графика нагрузок  
      food.fa$loadings[,2], ##по факторам  
      xlab = "Фактор 1",  
      ylab = "Фактор 2",  
      ylim = c(-1,1),  
      xlim = c(-1,1),  
      main = "Варимакс")  
text(food.fa.varimax$loadings[,1]-0.08, ##задаём подписи для графика  
      food.fa.varimax$loadings[,2]+0.08,  
      colnames(food),  
      col="blue")  
abline(h = 0, v = 0)  
abline(h = 0, v = 0) ##осевые линии для удобства
```

Рисунок 3 – Выполнение факторного анализа и его представление

```
Call:
factanal(x = food, factors = 2)

Uniquenesses:
      Oil Density Crispy Fracture Hardness
0.334  0.156    0.042  0.256    0.407

Loadings:
      Factor1 Factor2
oil      -0.816
Density  0.919
Crispy   -0.745  0.635
Fracture 0.645 -0.573
Hardness          0.764

      Factor1 Factor2
ss loadings  2.490  1.316
Proportion Var 0.498  0.263
Cumulative Var 0.498  0.761

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 0.27 on 1 degree of freedom.
The p-value is 0.603
```

Рисунок 4 – Результаты факторного анализа

Интерпретация результатов.

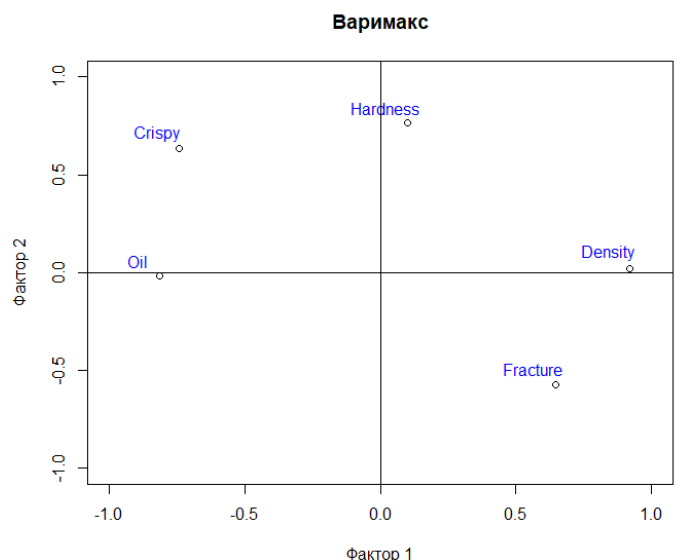


Рисунок 5 – График на основе полученных результатов анализа

Таким образом, необходимо отметить возможность применения только встроенных инструментов самого языка, скорость работы и простоту интерпретации полученных результатов за счёт их небольшого количества. Одновременно с этим, графическая интерпретация результатов заняла больший объём кода, при том, что графическое представление в данном языке в случае необходимости добавления дополнительных настроек, значительно усложнит поставленную задачу. Дополнительно, нет

возможности вывести дополнительные данные по результатам анализа, а настройка исходных параметров задачи не содержит большого количества необходимых компонентов, что упрощает подачу значений, но может и негативно сказаться на результате.

Пример проведения факторного анализа в Python.

Загрузка необходимых библиотек

```
In [1]: import pandas as pd # библиотека для работы с данными, в т.ч. чтение файлов формата CSV

In [3]: pip install factor_analyzer # установка Factor Analyzer

Defaulting to user installation because normal site-packages is not writeableNote: you may need to restart the kernel to use updated packages.

Requirement already satisfied: factor_analyzer in c:\users\астафьев рустам\appdata\roaming\python\python39\site-packages (0.4.1)
Requirement already satisfied: scipy in c:\programdata\anaconda3\lib\site-packages (from factor_analyzer) (1.9.1)
Requirement already satisfied: numpy in c:\programdata\anaconda3\lib\site-packages (from factor_analyzer) (1.21.5)
Requirement already satisfied: scikit-learn in c:\programdata\anaconda3\lib\site-packages (from factor_analyzer) (1.0.2)
Requirement already satisfied: pandas in c:\programdata\anaconda3\lib\site-packages (from factor_analyzer) (1.4.4)
Requirement already satisfied: pre-commit in c:\users\астафьев рустам\appdata\roaming\python\python39\site-packages (from factor_analyzer) (2.20.0)
Requirement already satisfied: pytz>=2020.1 in c:\programdata\anaconda3\lib\site-packages (from pandas->factor_analyzer) (2022.1)
Requirement already satisfied: python-dateutil>=2.8.1 in c:\programdata\anaconda3\lib\site-packages (from pandas->factor_analyzer) (2.8.2)
Requirement already satisfied: cfgv>=2.0.0 in c:\users\астафьев рустам\appdata\roaming\python\python39\site-packages (from pre-commit->factor_analyzer) (3.3.1)
Requirement already satisfied: virtualenv>=20.0.8 in c:\users\астафьев рустам\appdata\roaming\python\python39\site-packages (from pre-commit->factor_analyzer) (20.17.1)

In [4]: from factor_analyzer import FactorAnalyzer ##загружаем метод FactorAnalyzer
from factor_analyzer.factor_analyzer import calculate_bartlett_sphericity ##загружаем метод calculate_bartlett_sphericity
from factor_analyzer.factor_analyzer import calculate_kmo ##загружаем метод calculate_kmo
```

Рисунок 6 – Загрузка необходимых библиотек в Python

Загрузка и обработка файла до необходимого представления набора данных.

```
dftrain = pd.read_csv('C:/Users/Астафьев Рустам/Desktop/train.csv') # загружаем данные
dftest = pd.read_csv('C:/Users/Астафьев Рустам/Desktop/test.csv') # загружаем тестовые данные
df_ori = pd.concat([dftrain, dftest], sort=False) # объединяем тестовые данные и исходные для проверки загрузки
df = df_ori.copy() # загружаем необходимые данные оценок
df = df.iloc[:,8:24] # загружаем необходимые столбцы данных
df = df.dropna() # исключаем пропущенные значения
df # выводим набор данных
```

	Wi-Fi в полете	Удобство времени вылета/прибытия	Простота онлайн-бронирования	Расположение выхода на посадку	Еда и напитки	Онлайн-доска	Комфорт сиденья	Развлечения в полете	обслуживание на борту	Вкл. пространства для ног	Комфорт пространства для ног	Обработка багажа	Услуг регистрации
0	3	4	3	1	5	3	5	5	4	3	4		
1	3	2	3	3	1	3	1	1	1	5	3		
2	2	2	2	2	5	5	5	5	4	3	4		
3	2	5	5	5	2	2	2	2	2	5	3		
4	3	3	3	3	4	5	5	3	3	4	4		
...		
25971	3	3	3	1	4	3	4	4	3	2	4		
25972	4	4	4	4	4	4	4	4	4	5	5		
25973	2	5	1	5	2	1	2	2	4	3	4		
25974	3	3	3	3	4	4	4	4	3	2	5		
25975	2	5	2	5	4	2	2	1	1	2	1		

129487 rows x 16 columns

Рисунок 7 – Загрузка данных

Предварительный анализ данных.

```
calculate_bartlett_sphericity(df) # проверяем критерий Бартлетта на равенство дисперсий
(1100454.3463620122, 0.0)

kmo_all, kmo_model = calculate_kmo(df) # тест Кайзера-Мейера-Олкина (КМО)
print(kmo_model) # выводим результат КМО
0.7347314786301874

fa = FactorAnalyzer() # сокращаем команду для удобства написания
fa.fit(df) # выполняем метод на заданном наборе данных
ev, v = fa.get_eigenvalues() # находим собственный вектор коэффициентов
ev # отображаем результат

array([[3.80389837, 2.3721328, 2.16956004, 1.96182794, 1.06300589,
        0.94961228, 0.69607891, 0.53723377, 0.51374997, 0.46784182,
        0.36634408, 0.32884042, 0.29332142, 0.25443665, 0.18743275,
        0.03468288])
```

Рисунок 8 – Результаты предварительного анализа данных

Выполнение анализа.

```
fa = FactorAnalyzer(5, rotation='varimax') # на предыдущем шаге получилось значение 5, поэтому выполняем факторный анализ
fa.fit(df) # применяем его к набору данных # по 5 факторам
print(fa.loadings_) # печатаем полученную матрицу нагрузок

[[ 9.51222553e-02 1.34785546e-01 -9.00112205e-03 6.14101716e-01
  4.65372138e-01]
 [-9.57623178e-03 5.54625876e-02 -2.96640283e-04 5.89526399e-01
 -6.49140294e-03]
 [-3.22972069e-02 3.10800907e-02 -2.35280238e-03 7.72955089e-01
  4.48606156e-01]
 [ 1.25845034e-02 -4.67148473e-02 4.77284142e-03 6.82653004e-01
 -1.11332166e-01]
 [ 7.70829829e-01 4.10135156e-03 -1.80185150e-02 3.06650441e-02
 3.46802825e-02]
 [ 2.89549137e-01 1.22384915e-01 -9.53510983e-03 1.08246336e-01
 7.54004532e-01]
 [ 7.56387765e-01 7.95257130e-02 -1.38440023e-02 -2.64576486e-02
 2.09396962e-01]
 [ 7.67525559e-01 4.66055427e-01 -7.83337056e-02 4.09450715e-02
 2.32560625e-02]
 [ 8.52709956e-02 7.01341670e-01 -1.92813545e-02 1.03362679e-02
 4.71336117e-02]
 [ 5.78302900e-02 4.86147654e-01 2.34399005e-02 4.31284207e-02
 9.26339930e-02]
 [ 3.64249935e-02 7.64505664e-01 6.93860772e-03 4.62035697e-02
 -3.50870361e-02]
 [ 1.13416170e-01 2.87751429e-01 -1.30508044e-02 -2.70974643e-02
 1.33294580e-01]
 [ 3.57487778e-02 7.99371090e-01 -4.43771534e-02 4.63690079e-02
 -5.80222107e-02]
 [ 8.54194578e-01 8.49490650e-02 6.47275861e-04 -1.29069464e-03
 9.78446898e-02]
 [-1.56801143e-02 -1.42312276e-02 9.68663411e-01 9.12087649e-05
 -6.18584454e-03]
 [-1.73449745e-02 -1.94197429e-02 9.95885604e-01 -8.00343269e-04
 -8.27730962e-03]]
```

Рисунок 9 – Матрица нагрузок по итогам проведения факторного анализа

```
lmatrix = pd.DataFrame(fa.loadings_, index = list(df.columns), columns = ['Factor 1', 'Factor 2', 'Factor 3', 'Factor 4',
                                                                           'Factor 5'])
lmatrix # печатаем таблицу коэффициентов нагрузок
```

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Wi-Fi в полете	0.095122	0.134786	-0.009001	0.614102	0.465372
Удобство времени вылета/прибытия	-0.009570	0.055463	-0.002029	0.589520	-0.006491
Простота онлайн-бронирования	-0.032297	0.031080	-0.002353	0.772955	0.448806
Расположение выхода на посадку	0.012585	-0.046715	0.004773	0.682053	-0.111332
Еда и напитки	0.770830	0.004101	-0.018019	0.030955	0.034680
Онлайн-доска	0.269549	0.122385	-0.009535	0.108246	0.754005
Комфорт сиденья	0.756388	0.079520	-0.013844	-0.026458	0.209397
Развлечения в полете	0.767528	0.468055	-0.007833	0.040945	0.023256
Вкл.-обслуживание на борту	0.085271	0.701342	-0.019281	0.010336	0.047134
Комфорт пространства для ног	0.057830	0.486148	0.023440	0.043128	0.092634
Обработка багажа	0.036425	0.764506	0.006939	0.046204	-0.035087
Услуги регистрации	0.113410	0.287751	-0.013051	-0.027097	0.133295
Обслуживание на борту	0.035749	0.799371	-0.044377	0.046369	-0.058022
Чистота	0.854195	0.084949	0.000647	-0.001291	0.097845
Задержка вылета в минутах	-0.015880	-0.014231	0.968863	0.000091	-0.006188
Задержка прибытия в минутах	-0.017345	-0.019420	0.995888	-0.000800	-0.008277

Рисунок 10 – Матрица коэффициентов нагрузок для конечного набора факторов

Пример интерпретации результатов.

```
lmatrix.sort_values('Factor 1', ascending=False) # печатаем таблицу коэффициентов факторов по убыванию для фактора 1
```

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Чистота	0.854195	0.084949	0.000947	-0.001291	0.097845
Еда и напитки	0.770830	0.004101	-0.018019	0.030665	0.034680
Развлечения в полете	0.767526	0.466055	-0.007833	0.040945	0.023256
Комфорт сиденья	0.759388	0.079526	-0.013844	-0.026458	0.209397
Онлайн-доска	0.289549	0.122385	-0.009535	0.108248	0.754005
Услуги регистрации	0.113416	0.287751	-0.013051	-0.027097	0.133295
Wi-Fi в полете	0.095122	0.134786	-0.009001	0.614102	0.465372
Вкл. -обслуживание на борту	0.085271	0.701342	-0.019281	0.010336	0.047134
Комфорт пространства для ног	0.057830	0.486148	0.023440	0.043128	0.092834
Обработка багажа	0.036425	0.784506	0.006939	0.046204	-0.035087
Обслуживание на борту	0.035749	0.799371	-0.044377	0.046369	-0.059022
Расположение выхода на посадку	0.012585	-0.048715	0.004773	0.682853	-0.111332
Удобство времени вылета/прибытия	-0.009576	0.055483	-0.000297	0.589526	-0.008491
Задержка вылета в минутах	-0.015680	-0.014231	0.988963	0.000091	-0.008186
Задержка прибытия в минутах	-0.017345	-0.019420	0.995886	-0.000800	-0.008277
Процента онлайн-бронирования	-0.032297	0.031080	-0.002353	0.772955	0.448606

Рисунок 11 – Матрица коэффициентов для фактора №1

В конечном счёте, можно сделать вывод о том, что для использования данного метода необходимо владеть языком и его особенности для успешной подгрузки файлов, а также знать особенности использования библиотек Python, при том, что визуализация и конечная интерпретация имеет более наглядный характер, а конкретные настройки отображения, отличающиеся общей простотой, позволяют представлять результаты работы анализа в любом запрашиваемом виде. При этом, количество настраиваемых параметров анализа и большой стек данных о его результатах, в том числе промежуточных, позволяет гораздо тоньше настроить, как сам процесс анализа, так и конечное его представление.

Сравнительная таблица характеристик реализаций в R и в Python.

Опишем каждую из реализаций в разбивке по характеристикам, указанным в Таблице 1 и представим в сравнении в Таблице 2.

Таблица 2 – Сравнительная таблица реализаций по характеристикам

Характеристика	Реализация в R	Реализация в Python
Простота использования	Реализация в R отличается общей простотой, так как малое число параметров и применения количество методов позволяет быстро сориентироваться, как в использовании, так и в интерпретации результатов.	Реализация в Python простотой не отличается, но позволяет подробно использовать множество параметров, которые язык позволяет использовать по результатам проведения анализа непосредственно в языке.
Простота внедрения	R является специфическим языком, который не предполагает прямого программного внедрения его методов в конечные результаты и/или программы.	Python распространённый язык, внедрение которого возможно в пределах многих языков программирования или программных решений, в том числе в разных статистических пакетах.
Распространённость языка	R является распространённым языком в основном в среде статистического анализа данных.	Python является популярным языком для написания многих программных приложений, в том числе в среде анализа статистических и больших данных.
Возможности интерпретации	Результаты анализа формируют ограниченное число параметров, использование которых осложняется спецификой языка и применения их в контексте использования команд/методов.	Результаты проведённого анализа формируют обширное число возможных к использованию результатов, так как результаты интерпретируются в языке в форме классов, что формирует многокомпонентный набор данных, коэффициентов, которые могут быть использованы в дальнейшем исследовании.

<p>Возможности оценки результатов</p>	<p>Язык подразумевает встроенную оценку результатов исследования.</p>	<p>Язык включает в себя методы оценки результатов анализа, но их необходимо вызывать дополнительно.</p>
<p>Простота предварительного анализа данных</p>	<p>Для предварительного анализа данные необходимо готовить дополнительно, а также изучать применение и реализацию методов тестирования наборов данных.</p>	<p>Для предварительного анализа существует встроенная система тестов в класс факторного анализа, что упрощает использование и применение данных тестов к конечному набору исходных данных.</p>

Таким образом, можно сказать, что в Python факторный анализ раскрыт более широко, но именно эта широта усложняет его применения для непрофессионального анализа или его применения без конкретных знаний языка/статистических законов.

Алгоритм выбора реализации факторного анализа.

Синтезируем алгоритм выбора языка реализации факторного анализа на основе применения критериального подхода, то есть сформируем характеристики, дадим им комплексную оценку таким образом, чтобы исследователь мог оценить необходимость каждой характеристики, просуммировать результаты и получить предпочтительный языка реализации.

Переформулируем выбранные и оцененные характеристики в соответствии с необходимостью/возможностью использования.

Таблица 3 – Таблица характеристик, сформулированная в виде вопросов по анализу

Характеристика	Формулировка для алгоритма
Простота использования	Достаточно ли в качестве результатов анализа получить только матрицу весов факторов и общую оценку модели?
Простота внедрения	Необходимо ли внедрять текущее программное решение в более крупный проект исследования?
Распространённость языка	Знания какого языка Вы оцениваете, как наиболее полные, либо по какому языку в Вашем проекте есть специалист?
Возможности интерпретации	Достаточно ли для интерпретации результатов получить матрицу факторов и общей оценки модели?
Возможности оценки результатов	Есть ли у Вас достаточные знания по проведению статического анализа на предмет оценки полученных результатов?
Простота предварительного анализа данных	Есть ли у Вас достаточные знания по проведению статистического анализа на предмет предварительного анализа данных?

Теперь составим оценку характеристики в контексте необходимости применения конкретной реализации, влияющей на применения языка. Для этого характеристики поделим на 2 группы: напрямую влияет на выбор языка реализации и косвенно влияет на выбор языка реализации.

Таблица 4 – Деление характеристик по группам влияния на выбор языка реализации

Характеристика	Группа
Простота использования	косвенно влияет на выбор языка реализации
Простота внедрения	напрямую влияет на выбор языка реализации
Распространённость языка	напрямую влияет на выбор языка реализации

Возможности интерпретации	косвенно влияет на выбор языка реализации
Возможности оценки результатов	косвенно влияет на выбор языка реализации
Простота предварительного анализа данных	косвенно влияет на выбор языка реализации

Те характеристики, которые влияют напрямую, будем учитывать с весом, в 2 раза большим, чем факторам, косвенно влияющим на выбор:

Таблица 5 – Таблица оценки ответов по вопросам-характеристикам

Вопрос	Оценка ответов
Достаточно ли в качестве результатов анализа получить только матрицу весов факторов и общую оценку модели?	Диапазон от -5 до 5: -5: да, достаточно; 0: нет общей уверенности; 5: нет, необходимо ещё набор характеристик результатов.
Необходимо ли внедрять текущее программное решение в более крупный проект исследования?	Диапазон от -10 до 10: -10: решение внедрять никуда не нужно, в том числе, получить матрицу коэффициентов; 0: возможно, что решение необходимо будет применять в дальнейшем, что зависит от результатов анализа; 10: решение необходимо внедрять далее, применять дальнейшие методы исследования.
Знания какого языка Вы оцениваете, как наиболее полные, либо по какому языку в Вашем проекте есть специалист?	Диапазон от -10 до 10: -10: сильные знания только языка R; 0: умеренные знания обоих языков; 10: сильные знания только языка Python.
Достаточно ли для интерпретации результатов получить матрицу факторов и общей оценки модели?	Диапазон от -5 до 5: -5: да, достаточно; 0: нет общей уверенности; 5: нет, необходимо ещё набор характеристик результатов.
Есть ли у Вас достаточные знания по проведению	Диапазон от -5 до 5: -5: да, достаточно;

статического анализа на предмет оценки полученных результатов?	0: нет общей уверенности; 5: нет, необходимо ещё набор характеристик для оценки.
Есть ли у Вас достаточные знания по проведению статистического анализа на предмет предварительного анализа данных?	Диапазон от -5 до 5: -5: да, достаточно; 0: нет общей уверенности; 5: нет, необходимо ещё набор методов для предварительного анализа.

Все оценки необходимо просуммировать и, чем дальше от 0 сумма, тем большее предпочтение следует отдать реализации в Python, чем меньше 0, тем большее предпочтение следует отдать реализации в R. Конечное решение с таблицей ответов может послужить для принятия управляющего решения исследователю или руководителю проекта.

Приведём пример с ответами на вопросы для проведения анализа:

Таблица 6 – Пример применения оценки ответов на вопросы-характеристики

Вопрос	Ответ	Оценка
Достаточно ли в качестве результатов анализа получить только матрицу весов факторов и общую оценку модели?	Нет, недостаточно, необходимо выделить матрицу весов факторов, а на их основе уже получить сами факторы, пересчитать статистические характеристики для выделенных факторов.	3
Необходимо ли внедрять текущее программное решение в более крупный проект исследования?	Нет, необходимости нет, текущего решения достаточно для интерпретации, но необходимо.	-5
Знания какого языка Вы оцениваете, как наиболее полные, либо по какому языку в Вашем проекте есть специалист?	Язык Python известен и применяется лучшего всего, но есть знания и в R.	8
Достаточно ли для интерпретации результатов получить матрицу	Уверенности нет, так как многое зависит от самих результатов.	0

факторов и общей оценки модели?		
Есть ли у Вас достаточные знания по проведению статического анализа на предмет оценки полученных результатов?	Да, достаточные знания есть, но предпочтительно наличие встроенных методов дополнительной оценки результатов.	1
Есть ли у Вас достаточные знания по проведению статистического анализа на предмет предварительного анализа данных?	Данные подверглись предварительному анализу в другом программном пакете, поэтому они уже готовы к факторному анализу и являются факторизируемыми.	-2
	Сумма:	5

Таким образом, предпочтительнее выбрать язык Python для проведения факторного анализа в рамках текущего решения.

Таблица оценок может быть применима для выбора конкретного языка реализации, а алгоритм выбора может выглядеть следующим образом:

1. Выделить специалиста/исследователя, способного по поставленной задаче ответить на вопросы и поставить оценки в соответствии со шкалой критериев.
2. Провести оценку по шкале критериев в соответствии с таблицей вопросов.
3. Просуммировать оценки и принять решение о применении конкретного языка реализации.
4. Утвердить язык реализации у руководителя проекта/исследования.

Заключение

В результате проведения работы был выполнен теоретический обзор реализации методов факторного анализа в R и в Python, выделены основные необходимые характеристики методов, проведён сравнительный анализ на основе характеристик, по которым далее составлена таблица вопросов с оценками, на основе которых и был сформирован алгоритм выбора языка реализации. Был приведён пример ответов на вопросы и общий алгоритм

выбора реализации. Применимость конкретного метода может быть обоснована только внедрением его в последующие исследования, по итогам которых и возможна доработка метода по отдельным характеристикам.

Список источников

1. Габдракипова, Р. И. Применение факторного анализа в экономике / Р. И. Габдракипова, Р. Ф. Еркеев // В МИРЕ науки и ИННОВАЦИЙ: сборник статей международной научно-практической конференции: в 8 частях, Пермь, 25 декабря 2016 года. Том Часть 1. – Пермь: Общество с ограниченной ответственностью "Аэтерна", 2016. – С. 134-137. – EDN XHGХЕВ.
2. Саломатина, Т. В. Интегральные методы факторного анализа: перспективы использования в экономике: монография / Т. В. Саломатина; Т. В. Саломатина; М-во образования и науки Российской Федерации, Саратовский гос. социально-экономический ун-т. – Саратов: Саратовский гос. социально-экономический ун-т, 2011. – 147 с. – ISBN 978-5-4345-0065-4. – EDN QVEJDV.
3. Кулешов, И. Н. Факторный анализ инновационного развития экономики России / И. Н. Кулешов // Галактика науки–2021: Материалы Всероссийской (национальной) научно-практической конференции, Краснодар, 14–17 апреля 2021 года / Кубанский государственный университет. Том 1. – Краснодар: Кубанский государственный университет, 2021. – С. 117-121. – EDN OVNWRJ.
4. Factanal: Factor Analysis / [Электронный ресурс] // RDocumentation: [сайт]. — URL: <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/factanal> (дата обращения: 06.04.2023).
5. Factoextra: Extract and Visualize the Results of Multivariate Data Analyses / [Электронный ресурс] // RDocumentation: [сайт]. — URL:

<https://www.rdocumentation.org/packages/factoextra/versions/1.0.7> (дата обращения: 06.04.2023).

6. Factor-analyzer 0.4.1 / [Электронный ресурс] // Python Package Index: [сайт]. — URL: <https://pypi.org/project/factor-analyzer/> (дата обращения: 06.04.2023).

7. FactorAnalyzer documentation / [Электронный ресурс] // Read The Docs: [сайт]. — URL: <https://factor-analyzer.readthedocs.io/en/latest/index.html> (дата обращения: 06.04.2023).

8. Егошин, В. Л. Анализ главных компонент и факторный анализ в программной среде R / В. Л. Егошин, Н. В. Саввина, А. М. Гржибовский // West Kazakhstan Medical Journal. — 2020. — № 1(62). — С. 6-14. — EDN GFCMWD.

9. Андросова, Л. Н. О применении факторного и компонентного анализа в инвестировании предприятий / Л. Н. Андросова, Н. И. Москаленко // Вестник Белгородского университета потребительской кооперации. — 2008. — № 1(25). — С. 188-190. — EDN JWMYRF.

10. Демичев, В. В. Автоматизация статистического анализа факторов эффективности государственной поддержки сельского хозяйства с применением языка программирования Python / В. В. Демичев, А. А. Нестратова // Электронный сетевой политематический журнал "Научные труды КубГТУ". — 2020. — № 6. — С. 1-10. — EDN NQQZQM.

11. Прохоренков, П. А. Методы кластерного анализа в региональных исследованиях / П. А. Прохоренков, Т. В. Регер, Н. В. Гудкова // Фундаментальные исследования. — 2022. — № 3. — С. 100-106. — EDN KOVJWZ.

References

1. Gabdrakipova, R. I. Primenenie faktornogo analiza v e`konomike / R. I. Gabdrakipova, R. F. Erkeev // V MIRE nauki i INNOVACIJ: sbornik statej mezhdunarodnoj nauchno-prakticheskoj konferencii: v 8 chastyax, Perm`, 25

dekabrya 2016 goda. Tom Chast` 1. – Perm`: Obshhestvo s ogranichennoj otvetstvennost`yu "Ae`terna", 2016. – S. 134-137. – EDN XHGXEБ.

2. Salomatina, T. V. Integral`ny`e metody` faktornogo analiza: perspektivy` ispol`zovaniya v e`konomie: monografiya / T. V. Salomatina; T. V. Salomatina; M-vo obrazovaniya i nauki Rossijskoj Federacii, Saratovskij gos. social`no-e`konomicheskij un-t. – Saratov: Saratovskij gos. social`no-e`konomicheskij un-t, 2011. – 147 s. – ISBN 978-5-4345-0065-4. – EDN QVEJDV.

3. Kuleshov, I. N. Faktorny`j analiz innovacionnogo razvitiya e`konomiki Rossii / I. N. Kuleshov // Galaktika nauki–2021: Materialy` Vserossijskoj (nacional`noj) nauchno-prakticheskoy konferencii, Krasnodar, 14–17 aprelya 2021 goda / Kubanskij gosudarstvenny`j universitet. Tom 1. – Krasnodar: Kubanskij gosudarstvenny`j universitet, 2021. – S. 117-121. – EDN OVNWRJ.

4. Factanal: Factor Analysis / [E`lektronny`j resurs] // RDocumentation: [sajt]. — URL:

<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/factanal> (data obrashheniya: 06.04.2023).

5. Factoextra: Extract and Visualize the Results of Multivariate Data Analyses / [E`lektronny`j resurs] // RDocumentation: [sajt]. — URL: <https://www.rdocumentation.org/packages/factoextra/versions/1.0.7> (data obrashheniya: 06.04.2023).

6. Factor-analyzer 0.4.1 / [E`lektronny`j resurs] // Python Package Index: [sajt]. — URL: <https://pypi.org/project/factor-analyzer/> (data obrashheniya: 06.04.2023).

7. FactorAnalyzer documentation / [E`lektronny`j resurs] // Read The Docs: [sajt]. — URL: <https://factor-analyzer.readthedocs.io/en/latest/index.html> (data obrashheniya: 06.04.2023).

8. Egoshin, V. L. Analiz glavny`x komponent i faktorny`j analiz v programmnoj srede R / V. L. Egoshin, N. V. Savvina, A. M. Grzhibovskij // West Kazakhstan Medical Journal. – 2020. – № 1(62). – S. 6-14. – EDN GFCMWD.

9. Androsova, L. N. O primeneniі faktornogo i komponentnogo analiza v investirovanii predpriyatij / L. N. Androsova, N. I. Moskalenko // Vestnik Belgorodskogo universiteta potrebitel'skoj kooperacii. – 2008. – № 1(25). – S. 188-190. – EDN JWMYRF.

10. Demichev, V. V. Avtomatizaciya statisticheskogo analiza faktorov e'ffektivnosti gosudarstvennoj podderzhki sel'skogo khozyajstva s primeneniem yazyka programmirovaniya Python / V. V. Demichev, A. A. Nestratova // E'lektronnyj setевой politematicheskij zhurnal "Nauchny'e trudy` KubGTU". – 2020. – № 6. – S. 1-10. – EDN NQQZQM.

11. Proxorenkov, P. A. Metody` klasternogo analiza v regional'ny`x issledovaniyax / P. A. Proxorenkov, T. V. Reger, N. V. Gudkova // Fundamental'ny'e issledovaniya. – 2022. – № 3. – S. 100-106. – EDN KOVJWZ.

Для цитирования: Астафьев Р.У., Пронина Е.В., Пихтилькова О.А., Параскевопуло О.Р., Морозова Т.А., Кузнецова Е.Ю., Немировская-Дутчак О.Э. Синтез алгоритма выбора реализации методов факторного анализа эконометрических данных в R и Python // Московский экономический журнал. 2023. № 4. URL: <https://qje.su/ekonomicheskaya-teoriya/moskovskij-ekonomicheskij-zhurnal-4-2023-41/>

© Астафьев Р.У., Пронина Е.В., Пихтилькова О.А., Параскевопуло О.Р., Морозова Т.А., Кузнецова Е.Ю., Немировская-Дутчак О.Э., 2023. Московский экономический журнал, 2023, № 4.