Факторный анализ

Многомерный анализ

Математические основы, разобранный пример, работа с инструментом

Русский

Вольфганг Бёме

wolfgang.boehme@economic-modeling.de

www.economic-modeling.de

8 ноября 2026 г.

Документация на этом сайте разработана при поддержке ИИ Claude (Anthropic).

Содержание

1 Введение и области применения **2**

2 Математические основы **2**
2.1 Факторная модель . . . 2
2.2 Коммунальность и уникальность . . . 2
2.3 Воспроизведённая матрица корреляций . . . 2
2.4 Метод главных осей (PAF) . . . 3
2.5 Мера KMO (Kaiser-Meyer-Olkin) . . . 3
2.6 Тест Бартлетта на сферичность . . . 3
2.7 Вращение Varimax . . . 3

3 Разобранный пример: балансовые показатели **4**
3.1 Исходные данные ($n = 24$, $p = 6$) . . . 4
3.2 Шаг 1: Матрица корреляций, KMO и тест Бартлетта . . . 4
3.3 Шаг 2: Собственные значения (критерий Кайзера) . . . 4
3.4 Шаг 3: Экстракция PAF . . . 5

4 Работа с интерактивным инструментом факторного анализа **5**
4.1 Ввод данных . . . 5
4.2 Выбор параметров . . . 6
4.3 Вкладки результатов . . . 6
4.4 Типичный рабочий процесс . . . 6

5 Сводка формул **7**

6 Рекомендуемая литература **7**

Введение и области применения

Факторный анализ — это метод выявления *скрытых факторов*, лежащих в основе наблюдаемых переменных. В отличие от МГК, который лишь максимизирует дисперсию, факторный анализ исходит из явной измерительной модели: каждая переменная моделируется как линейная комбинация общих факторов и переменно-специфической уникальности.

Типичные области применения:

- Построение шкал в психологии и социальных науках
- Валидация пунктов анкет (например, Большая пятёрка)
- Сокращение индикаторов при построении индексов (например, качество жизни)
- Выявление скрытых измерений в данных маркетинговых исследований

Математические основы

Факторная модель

Для p стандартизованных переменных $\mathbf{z} = (z_1, \ldots, z_p)^\top$ выполняется:

$$z_j = l_{j1}F_1 + l_{j2}F_2 + \cdots + l_{jk}F_k + u_j, \qquad j = 1, \ldots, p \tag{1}$$

где F_k — общие факторы ($\mathrm{Var}(F_k) = 1$, $\mathrm{Cov}(F_k, F_m) = 0$), l_{jk} — факторные нагрузки и u_j — специфическая дисперсия (уникальность).

В матричной записи:

$$\mathbf{Z} = \mathbf{F}\mathbf{L}^\top + \mathbf{U} \tag{2}$$

Коммунальность и уникальность

$$h_j^2 = \sum_{k=1}^{K} l_{jk}^2, \qquad u_j = 1 - h_j^2 \tag{3}$$

Коммунальность h_j^2 — доля дисперсии j-й переменной, объяснённой K факторами; u_j — необъяснённый остаток.

Воспроизведённая матрица корреляций

Из факторной модели следует:

$$\mathbf{R} = \mathbf{L}\mathbf{L}^\top + \mathbf{\Psi}, \qquad \mathbf{\Psi} = \mathrm{diag}(u_1, \ldots, u_p) \tag{4}$$

Цель экстракции: минимизировать остатки $\mathbf{R} - \hat{\mathbf{R}}$.

Метод главных осей (PAF)

Метод **Principal Axis Factoring** работает итеративно:

1. Начальная оценка коммунальностей: $h_j^{(0)} = \max_{m \neq j} |r_{jm}|^2$
2. Редуцированная матрица корреляций: $\mathbf{R}^* = \mathbf{R}$ с r_{jj}, замененными на h_j^2
3. Разложение $\mathbf{R}^*$ по собственным значениям; экстракция K наибольших положительных собственных значений и векторов
4. Обновление коммунальностей: $h_j^{(t+1)} = \sum_{k=1}^{K} l_{jk}^2$
5. Повторять до сходимости ($|h_j^{(t+1)} - h_j^{(t)}| < 10^{-6}$)

Мера KMO (Kaiser-Meyer-Olkin)

Мера KMO оценивает пригодность корреляционной структуры для факторного анализа:

$$\text{KMO} = \frac{\sum_{j \neq m} r_{jm}^2}{\sum_{j \neq m} r_{jm}^2 + \sum_{j \neq m} p_{jm}^2} \tag{5}$$

где p_{jm} — частная корреляция между z_j и z_m.

Таблица 1: Оценка меры KMO

Значение KMO	Оценка
≥ 0.90	превосходно
$0.80 - 0.89$	хорошо
$0.70 - 0.79$	удовлетворительно
$0.60 - 0.69$	посредственно
< 0.60	неприемлемо

Тест Бартлетта на сферичность

H_0: Матрица корреляций является единичной ($\mathbf{R} = \mathbf{I}$).

$$\chi^2 \approx -\left(n - 1 - \frac{2p + 5}{6}\right) \ln |\mathbf{R}|, \quad \text{df} = \frac{p(p-1)}{2} \tag{6}$$

Значимый результат ($p < 0.05$) указывает, что матрица корреляций отличается от единичной — необходимое условие для ФА.

Вращение Varimax

Идентично МГК (см. Руководство 1). Цель: простая структура — каждая переменная нагружает по возможности только один фактор.

Разобранный пример: балансовые показатели

Исходные данные ($n = 24$, $p = 6$)

Используется тот же синтетический набор данных, что и в Руководстве 1 (полная таблица данных приведена там): 24 предприятия с тремя показателями устойчивости (X_1 доля собственного капитала, X_2 ликвидность 2-й степени, X_3 покрытие внеоборотных активов II) и тремя показателями доходности (X_4 рентабельность продаж, X_5 рентабельность активов, X_6 рост выручки), все в процентах. Набор данных загружен в инструменте по умолчанию.

Цель факторного анализа: выявить латентные факторы, стоящие за показателями. В отличие от МГК, факторный анализ объясняет только *общую* дисперсию показателей; специфический остаток входит в уникальность u_j. Настройки: PAF, критерий Кайзера, Varimax (по умолчанию).

Шаг 1: Матрица корреляций, KMO и тест Бартлетта

Матрица корреляций (см. Руководство 1, шаг 2) содержит два блока с корреляциями 0.70–0.79 внутри и не более 0.31 между блоками.

Таблица 2: Критерии качества

Критерий	Значение	Оценка
KMO	0.678	посредственно (порог 0.6 выполнен)
Бартлетт χ^2	80.68 (df = 15)	$p < 0.001$, значимо

Расчёт по Бартлетту: $|\mathbf{R}| = 0.0183$, следовательно

$$\chi^2 = -\left(24 - 1 - \tfrac{2\cdot 6+5}{6}\right)\ln 0.0183 = -20.167 \cdot (-4.001) = 80.68, \qquad \mathrm{df} = \tfrac{6\cdot 5}{2} = 15.$$

Матрица корреляций высокозначимо отличается от единичной. Значение KMO лишь посредственное: каждый блок состоит всего из трёх показателей, поэтому частные корреляции внутри блоков остаются заметными. Минимальное требование KMO ≥ 0.6 тем не менее выполнено — факторный анализ допустим.

Шаг 2: Собственные значения (критерий Кайзера)

Таблица 3: Собственные значения матрицы корреляций

Фактор	Собств. знач.	Доля дисп.	Накоплен.
F1	2.971	49.52 %	49.52 %
F2	2.010	33.50 %	83.01 %
F3	0.397	6.61 %	89.62 %
F4–F6	< 0.3	–	–

Критерий Кайзера: **2 фактора** ($\lambda \geq 1$).

Шаг 3: Экстракция PAF

Начальные коммунальности (наибольшая квадрированная корреляция в строке): $h_1^{(0)} = 0.743^2 = 0.553$, $h_2^{(0)} = h_3^{(0)} = 0.771^2 = 0.595$, $h_4^{(0)} = h_6^{(0)} = 0.793^2 = 0.628$, $h_5^{(0)} = 0.727^2 = 0.529$. Диагональ $\mathbf{R}$ заменяется этими значениями, редуцированная матрица раскладывается, и процедура повторяется до сходимости.

Таблица 4: Матрица нагрузок PAF: без вращения и после вращения Varimax

	без вращения		Varimax			
Переменная	F1	F2	F1	F2	h_j^2	u_j
X_1 Доля СК	0.673	−0.497	**0.827**	0.122	0.699	0.301
X_2 Ликвидность	0.683	−0.522	**0.853**	0.112	0.739	0.261
X_3 Покрытие активов	0.664	−0.595	**0.890**	0.047	0.795	0.205
X_4 Рент. продаж	0.741	0.486	0.183	**0.867**	0.785	0.215
X_5 Рент. активов	0.618	0.506	0.082	**0.795**	0.638	0.362
X_6 Рост выручки	0.654	0.631	0.019	**0.909**	0.827	0.173
Объясн. дисп.	45.3 %	29.4 %	37.4 %	37.3 %	Σ 74.7 %	

Пример: $h_1^2 = 0.827^2 + 0.122^2 = 0.699$, $u_1 = 1 - 0.699 = 0.301$. Вместе факторы объясняют 74.7 % общей дисперсии — меньше, чем две главные компоненты МГК (83.0 %), поскольку PAF моделирует только общую дисперсию.

Воспроизведённая корреляция: факторная модель воспроизводит, например, $\hat{r}_{12} = 0.827 \cdot 0.853 + 0.122 \cdot 0.112 = 0.719$ (наблюдаемая: 0.710, остаток −0.009) и $\hat{r}_{14} = 0.827 \cdot 0.183 + 0.122 \cdot 0.867 = 0.257$ (наблюдаемая: 0.311). Малые остатки подтверждают двухфакторное решение.

Интерпретация:

- **Фактор 1 — финансовая устойчивость:** доля собственного капитала, ликвидность и покрытие активов (нагрузки 0.83–0.89). Он измеряет капитальную базу и устойчивость предприятия к кризисам.
- **Фактор 2 — доходность:** рентабельность продаж, рентабельность активов и рост выручки (нагрузки 0.80–0.91). Он измеряет способность получать прибыль и расти.
- Все перекрёстные нагрузки ниже 0.2: решение Varimax очень хорошо удовлетворяет критерию простой структуры. Наименьшая коммунальность у рентабельности активов (0.638) — она хуже всего объясняется факторами, но заметно превышает порог 0.4.

Работа с интерактивным инструментом факторного анализа

Ввод данных

Аналогично МГК (Вручную, CSV, Пример). По умолчанию загружены балансовые показатели 24 предприятий ($n = 24$, $p = 6$). Дополнительные примеры: удовлетворённость трудом ($n = 20$, $p = 6$), шкала личности (Большая пятёрка,

$n = 25$, $p = 8$) и поведение потребителей ($n = 18$, $p = 5$). Для каждого примера над таблицей ввода появляется пояснительный блок (данные, цель анализа, ожидаемый результат); он исчезает при нажатии «Очистить» или загрузке CSV-файла.

Выбор параметров

Метод экстракции
Метод главных осей (PAF): рекомендуется; корректно моделирует скрытые факторы через итеративную оценку коммунальностей. *Метод главных компонент:* быстрый, но объясняет также специфическую дисперсию.

Количество факторов
Автоматически: критерий Кайзера для матрицы корреляций. Или вручную задать 1–5 факторов.

Вращение
Varimax (стандарт): ортогональное вращение; факторы остаются некоррелированными. *Без вращения:* невращённое решение для прямого сравнения с МГК.

Порог нагрузок
Выделение при $|l| \geq 0.4$ (настраивается). Значения ≥ 0.5 считаются существенными.

KMO и Бартлетт
Флажок активирует расчёт проверки предпосылок.

Вкладки результатов

Обзор
Показатели + таблица собственных значений с маркером Кайзера.

Нагрузки
Матрица факторных нагрузок с цветовым выделением, коммунальности и уникальность по переменным.

Качество
Индикатор KMO, тест Бартлетта с p-значением, таблица коммунальностей с оценкой качества.

Графики
График осыпи, столбчатая диаграмма дисперсии с накопленной линией, график нагрузок (F1 vs. F2) с подписями переменных.

Интерпретация
Автоматический текст оценки с KMO, Бартлеттом и указанием о вращении.

Экспорт
CSV: нагрузки, коммунальности, собственные значения. PNG: график осыпи, нагрузок.

Типичный рабочий процесс

1. Загрузить данные и выбрать метод PAF

2. Оставить флажок KMO включённым
3. Нажать »**Вычислить факторный анализ**«
4. Вкладка «Качество»: проверить KMO ≥ 0.6 и значимость Бартлетта
5. График осыпи: при необходимости скорректировать число факторов вручную
6. Вкладка «Нагрузки»: содержательная интерпретация факторов
7. Переменные с $h_j^2 < 0.4$ при необходимости удалить и пересчитать

Важные замечания:

- Требуется минимум $n = 5$ полных наблюдений.
- Практическое правило: $n \geq 10 \cdot p$ для надёжных результатов.
- Переменные с $h_j^2 < 0.3$ плохо объяснены — рассмотреть их исключение.
- При перекрёстных нагрузках ($|l_{jk}| \geq 0.3$ на нескольких факторах): рассмотреть косоугольное вращение (Oblimin).

Сводка формул

Таблица 5: Формулы факторного анализа

Обозначение	**Формула**
Факторная модель	$z_j = \sum_k l_{jk} F_k + u_j$
Воспроизв. корреляция	$\mathbf{R} = \mathbf{L}\mathbf{L}^\top + \mathbf{\Psi}$
Коммунальность	$h_j^2 = \sum_k l_{jk}^2$
Уникальность	$u_j = 1 - h_j^2$
Мера KMO	$\text{KMO} = \sum r_{jm}^2 / (\sum r_{jm}^2 + \sum p_{jm}^2)$
Бартлетт χ^2	$-(n - 1 - \frac{2p+5}{6}) \ln \lvert\mathbf{R}\rvert$, $\text{df} = p(p-1)/2$

Рекомендуемая литература

- Comrey, A. L. & Lee, H. B. (1992). *A First Course in Factor Analysis*. LEA.
- Backhaus, K. et al. (2021). *Multivariate Analysemethoden*, 16. Aufl. Springer Gabler.
- Fabrigar, L. R. et al. (1999). Evaluating the use of exploratory FA in research. *Psychological Methods*, 4(3), 272–299.