Факторный анализ

ВРАЩЕНИЕ

ПРОЦЕДУРА ФАКТОРНОГО АНАЛИЗА

- Подготовка исходной матрицы
  - Очистка данных
  - Обработка выбросов
  - Стандартизация
- ВЫБОР факторной модели
- ИЗВЛЕЧЕНИЕ компонент/факторов (факторизация)
- **ВРАЩЕНИЕ факторов**
- Оценка факторных значений и ИНТЕРПРЕТАЦИЯ факторов

ШАГ 3

ВРАЩЕНИЕ

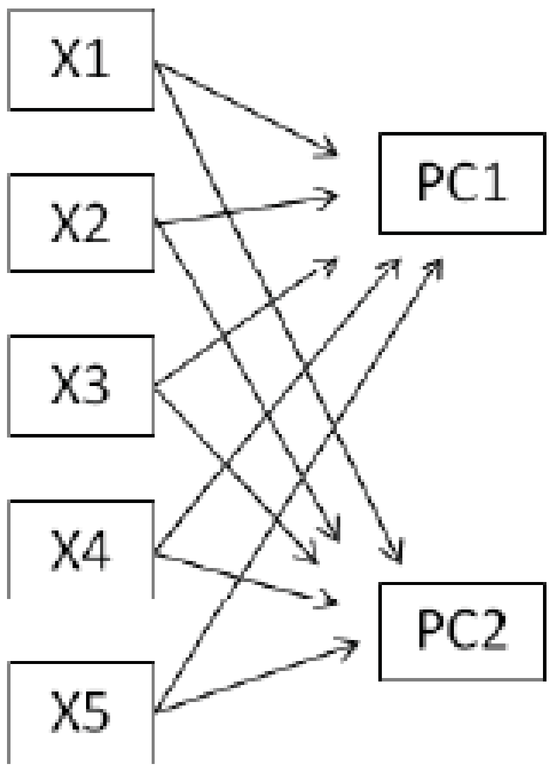
- **Факторизация** - Первое факторное решение
  - *Метод главных компонент*
  - *Общий факторный анализ*
  - *Факторизация максимального правдоподобия (MLF)*
  - *Факторизация не взвешенных наименьших квадратов (ULS)*
  - *Факторизация обобщенных наименьших квадратов (GLS)*
  - *Альфа факторизация*

- **Вращение** – Второе факторное решение
  - *Varimax, Oblimin и другие*

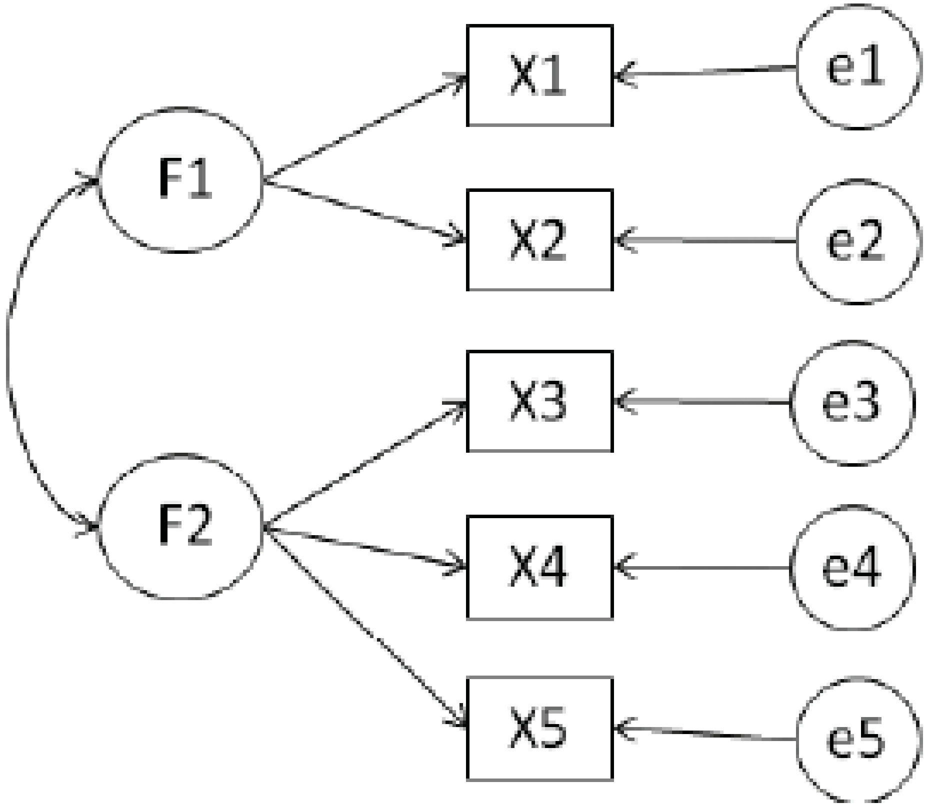
- ФАКТОРЫ и ФАКТОРНЫЕ НАГРУЗКИ

- Интерпретация

- Обычно факторы, полученные методом главных компонент, не поддаются достаточно наглядной интерпретации, поскольку переменные склонны нагружаться на множество факторов.
- Поэтому следующим шагом факторного анализа служит преобразование (**вращение**) факторов таким образом, чтобы облегчить их интерпретацию.



(a) Модель анализа главных компонент



(b) Модель факторного анализа

Факторный анализ - проблемы вращения и интерпретации

- Методы получения первичного факторного решения (метод главных факторов, метод минимальных остатков, центроидный метод, метод максимального правдоподобия и т.д.) не всегда позволяют производить интерпретацию факторной структуры.
- НАПРИМЕР:
- Метод главных факторов имеет высоконагруженный исходными показателями первый фактор, включающий в себя максимум разброса значений переменных, вычисляемых в проекции на первый фактор.
- Метод минимальных остатков имеет непредсказуемый характер, зависимый от начального приближения, что объясняется его критерием минимизации невязок восстанавливаемой и исходной корреляционной матрицей.
- Проблема факторного вращения связана с неоднозначностью факторных решений. В рамках критерия минимальных остатков удовлетворительное решение может иметь различный вид.

Методы вращения

- Для выделения лучшего решения используют различные **вращения**. Получаемое в результате вращения факторное решение должно обладать хорошей интерпретируемостью, суть которого заключается **в однозначном отнесение каждой исходной переменной лишь к одному из факторов**.
- Результирующие факторы могут быть как *ортогональны* друг другу, так и *косоугольны*. Косоугольное вращение приводит, как правило, к лучшему разнесению переменных на факторы в связи с большей возможностью выбора направления факторных осей и получения простой факторной структуры, когда исходные переменные максимально прижаты к факторным осям.

Имеется первичное факторное решение.

С помощью **вращения** необходимо получить матрицу простой структуры. Критерии основаны на двух соображениях:
а) необходимо определить **признаки простой структуры**
б) необходимо **выявить условия**, при которых простая структура выделяется однозначно и объективно.

Сами признаки простой структуры однозначно не установлены, поэтому, обычно, опираются на пункт б), то есть устанавливают некоторые **условия**.
Определить минимальные требования к простой структуре трудно, но если взять число факторов r и число переменных n, то всегда можно сказать, какая структура наиболее простая.

Факторная структура является наипростейшей, когда все переменные имеют факторную сложность, равную 1, т. е. когда каждая переменная имеет ненулевую нагрузку только на один общий фактор.

Мера факторной сложности переменной - дисперсия квадратов **факторных нагрузок переменной**.

$$\text{Факторная сложность переменной} = \frac{1}{r} \sum_{j=1}^{r} \left(b_{ij}^2 - \overline{b}_{ij}^2\right)^2,$$

где r — число столбцов факторной матрицы; b_{ij} — факторная нагрузка j-го фактора на i-ю переменную; $\overline{b}_{ij}$ — среднее значение квадратов факторных нагрузок в i-й строке.

Общности переменных исходной факторной структуры:

$$\sum_{j=1}^{r} b_{ij}^2 = h_i^2$$

- Если число факторов два и больше, то это означает, что в наиболее простой матрице факторной структуры:
- во-первых, каждая строка будет содержать только один ненулевой элемент,
- во-вторых, каждый столбец будет иметь несколько нулей и,
- в-третьих, для каждой пары столбцов нулевые элементы не совпадают.

	F1	*F2*	*F3*	*F4*	*F5*
Вес	0,346	-0,464	0,236	0,153	**0,602**
ИМТ	0,327	-0,531	0,207	0,182	**0,561**
ЧД	0,213	-0,418	0,125	**-0,711**	-0,055
С	-0,132	0,217	**-0,809**	-0,191	0,319
Л	0,075	-0,218	**0,785**	0,235	-0,364
КСР	**0,937**	-0,187	-0,212	0,075	-0,146
КСО	**0,92**	-0,171	-0,228	0,074	-0,143
КДР	**0,971**	0,167	0,011	0,021	-0,018
КДО	**0,979**	0,128	-0,032	0,009	-0,012
УО	**0,854**	0,372	0,178	-0,065	0,107

Первичное факторное решение

	F1	*F2*	*F3*	*F4*	*F5*
Вес	0,1349	-0,0739	0,0789	-0,0699	**0,8613**
ИМТ	0,0842	-0,1472	0,0952	-0,0701	**0,8584**
ЧД	0,086	-0,0774	0,0433	**-0,8493**	0,0873
С	-0,07	-0,052	**-0,9135**	0,0598	-0,1075
Л	0,0117	0,0282	**0,9229**	-0,0162	0,0647
КСР	**0,7393**	-0,6377	0,0145	-0,0933	0,151
КСО	**0,7295**	-0,6287	-0,0047	-0,0835	0,1379
КДР	**0,943**	-0,2564	0,0485	-0,0129	0,1191
КДО	**0,9316**	-0,2985	0,0165	-0,035	0,1329
УО	**0,9465**	0,0697	0,0524	-0,0015	0,1075

Вторичное факторное решение

- Факторное вращение осуществляется посредством умножения первичного факторного отображения A на матрицу преобразования V.
- Матрица преобразования определяется в соответствии с определённым критерием на элементы конечной факторной структуры. Минимизация или максимизация этого критерия позволяет найти оптимальную матрицу преобразования, доставляющую минимум или максимум целевой функции критерия.

ВРАЩЕНИЕ (источник Шовин В.А)

Математическая постановка задачи

Матрица $\underset{m \times g}{A \leftrightarrow} a_{ij}$ — матрица первичного факторного отображения размерности $m \times g$ весовых коэффициентов. Где m — число изучаемых параметров, g — число общих факторов.

Вращение заключается в следующей матричной операции: $V = A\Lambda$,

$\underset{m \times g}{V \leftrightarrow} v_{ij}$ — ~~косоугольная~~ факторная структура; Вторичная факторная структура

$\underset{g \times g}{\Lambda \leftrightarrow} \lambda_{ij}$ — матрица вращения.

• Дополнительные условия

Матрица корреляций $\underset{g\times g}{C} \leftrightarrow c_{ij}$ размерности $g \times g$ между конечными факторами, когда исходные данные стандартизированы (дисперсии переменных равны 1, а средние 0) вычисляется по формуле: $C = \Lambda^T \Lambda$.

Для матрицы вращения Λ должны выполняться соотношения:
$c_{ii} = \sum_{i=1}^{g} \lambda_{ij}^2 = 1$ (j=1,...,g).
А также ограничения типа неравенств:

$$|c_{ij}| = \left| \sum_{k=1}^{g} \lambda_{ki} \lambda_{kj} \right| \leq 1.$$

Если результирующее факторное решение должно быть ортогональным, то ограничения типа неравенств, заменяют ограничениями $c_{ij} = \underline{\sum_{k=1}^{g} \lambda_{ki} \lambda_{kj} = 0}.$

ВРАЩЕНИЕ

Задача факторного вращения соответствует максимизации или минимизации определённого критерия K, как функции от элементов матрицы результирующей факторной структуры:

$K = f(\{v_{ij}\}) = f(\{a_{ij}\}, \{\lambda_{ij}\})$. Поскольку элементы $\{a_{ij}\}$ заданы, то задача сводится к нахождению экстремума функции $K = f(\{\lambda_{ij}\})$ от независимых переменных $\{\lambda_{ij}\}$ с ограничениями:

$\sum_{i=1}^{g} \lambda_{ij}^2 = 1$ и

$\left|\sum_{k=1}^{g} \lambda_{ki}\lambda_{kj}\right| \leq 1$ или $\sum_{k=1}^{g} \lambda_{ki}\lambda_{kj} = 0$.

Также предлагается использовать следующие ограничения на вид результирующей факторной структуры. Общности переменных конечной факторной структуры должны быть не больше общностей переменных исходной факторной структуры, а также не меньше определенного порога значимости:

$h_i^v = \sqrt{\sum_{k=1}^{g} v_{ik}^2} \leq \sqrt{\sum_{k=1}^{g} a_{ik}^2} = h_i^a$ и $h_i^v \geq p$.

Критерии вращения

Квартимакс

$$K = \sum_{i=1}^{m} \sum_{p=1}^{g} v_{ip}^4.$$

Максимизация данного критерия приводит теоретически к минимальной сложности каждого исходного параметра равной 1, когда исходный параметр выражается только через один фактор

Варимакс

$$K = n \sum_{p=1}^{g} \sum_{i=1}^{m} \left(\frac{v_{ip}}{h_i} \right)^2 - \sum_{p=1}^{g} \left(\sum_{i=1}^{m} \frac{v_{ip}^2}{h_i^2} \right)^2. \qquad / m^2$$

Максимизация варимакс критерия соответствует максимизации дисперсий квадратов нагрузок факторов. Тем самым теоретическая сложность фактора уменьшается, нагрузки фактора близки к 0 или 1, и фактор можно наилучшим образом проинтерпретировать. Нормализация факторных нагрузок в данном критерии устраняет различие между вкладами отдельных параметров пропорциональное их общностям.

Обычно применяют нормированные факторные нагрузки, чтобы избавиться
от нежелательного влияния на результат вращения переменных с большой общностью.

Здесь h_i^2 - общность i-той переменной первичного факторного решения.

Облимакс

$$K = \frac{\sum_{i=1}^{m} \sum_{p=1}^{g} v_{ip}^4}{\left(\sum_{i=1}^{m} \sum_{p=1}^{g} v_{ip}^2\right)^2}.$$

Максимизация данного критерия соответствует максимизации эксцесса случайной величины ξ, представленной выборкой v_{ij} и $-v_{ij}$. В результате максимизируется доля больших и маленьких (близких к нулю) элементов факторной структуры.

Облимин

$$K = \sum_{p<q=1}^{g} \left[n \sum_{i=1}^{m} \left(\frac{v_{ip}}{h_i}\right)^2 \left(\frac{v_{iq}}{h_i}\right)^2 - \gamma \left(\sum_{i=1}^{m} \frac{v_{ip}^2}{h_i^2}\right) \left(\sum_{i=1}^{m} \frac{v_{iq}^2}{h_i^2}\right) \right],$$

При $\gamma = 0$ — это нормализованный квартимин критерий,
при $\gamma = 1$ критерий называется коваримин,
при $\gamma = 0.5$ критерий называется биквартимин.

Минимизация коваримин-критерия соответствует минимизации ковариации между квадратами элементов различных пар факторов конечной факторной структуры. Коваримин-критерий теоретически даёт ортогональное решение.

Факторный анализ в R

- Для проведения факторного анализа в R можно использовать различные возможности:
- Встроенная функция.

  Функция **princomp**, реализует метод главных компонент.
- Пакет многомерного анализа **mva**.

  Функция **factanal** – общий факторный анализ.
- Факторный анализ с помощью пакета **psych**, см. ниже, Таблица 14.1.

  Функция **principal()** – метод главных компонент без вращения, и метод главных компонент с вращением.

  Функция **fa()** – факторный анализ.

Таблица 14.1. Полезные функции для проведения факторного анализа при помощи пакета `psych`

Функции	Описание
`principal()`	Анализ главных компонент с возможностью поворота осей*
`fa()`	Факторный анализ методом главных осей, минимальных остатков, взвешенных наименьших квадратов или наибольшего правдоподобия
`fa.parallel()`	График собственных значений (scree plot) с параллельным анализом
`factor.plot()`	Графическое изображение результатов факторного анализа или анализа главных компонент
`fa.diagram()`	Графическое изображения матриц нагрузок факторного анализа или анализа главных компонент
`scree()`	График собственных значений для факторного анализа и анализа главных компонент

* Это, по сути, процедура EFA, проводимого с первоначальным построением осей-факторов по методу PCA. Поскольку данный способ построения системы ортогональных осей, описывающих экспериментальные данные, наиболее обоснован математически, он вынесен в отдельную процедуру. – *Прим. пер.*

Источник: Kabacoff. R in Action.

Две основные настройки параметров в факторном анализе fa(): метод Факторизации и метод Вращения.

- 1) Методы факторизации **fm**
- fm="minres" will do a minimum residual as will fm="**uls**". Both of these use a first derivative, *не взвешенные наименьшие квадраты*
- fm="ols" differs very slightly from "minres" in that it minimizes the entire residual matrix using an OLS procedure but uses the empirical first derivative.
- fm="wls" will do a weighted least squares (WLS) solution,
- fm="gls" does a generalized weighted least squares (**GLS**), *обобщенные наименьшие квадраты*
- fm="pa" will do the principal factor solution, **главные компоненты**
- fm="ml" will do a maximum likelihood factor analysis, **максимальное правдоподобие**
- fm="minchi" will minimize the sample size weighted chi square when treating pairwise correlations with different number of subjects per pair. fm ="minrank" will do a minimum rank factor analysis.
- fm="alpha" will do alpha factor analysis.

- 2) Методы вращения **rotate**
- none", "varimax", "quartimax", "bentlerT", "equamax", "varimin", "geominT" and "bifactor" - обеспечивают ортогональное вращение.

- "promax", "oblimin", "simplimax", "bentlerQ, "geominQ" and "biquartimin" and "cluster" - обеспечивают косоугольное вращение.

- По умолчанию используется "oblimin".

Оригинальное решение в анализе главных компонент максимизирует сумму квадратов факторных нагрузок

- *Вращение варимакс* [varimax rotation] является ортогональным вращением факторных осей с целью максимизировать дисперсию квадратов нагрузок фактора (колонки) по всем переменным (рядам) в факторной матрице, **Решение варимакс обеспечивает результаты, которые позволяют настолько, насколько это возможно, связать каждую переменную с каким-то одним фактором.** Это наиболее распространенный метод вращения.
- *Вращение квартимакс* [quartimax rotation] является ортогональной альтернативой, которая **минимизирует число факторов, необходимых для объяснения каждой переменной**. Этот тип вращения часто генерирует общий фактор, на который большинство переменных нагружается в высокой или умеренной степени. Такая факторная структура обычно не оказывается полезной для исследовательских целей.
- *Вращение эквимакс* [equimax rotation] является компромиссом между критериями варимакс и квартимакс.

- *Прямое вращение облимин* [direct oblimin rotation] является стандартным методом, когда исследователь стремится к неортогональному (косоугольному) решению – то есть, решению, в котором факторам разрешено коррелировать. Это будет приводить к более высоким собственным значениям, но уменьшать возможности интерпретации факторов.
- *Вращение промакс* [promax rotation] является альтернативным неортогональным (косоугольным) методом вращения, который в вычислительном отношении реализуется быстрее, чем метод прямого облимина, и поэтому его иногда используют при очень крупных массивах данных.

ПРИМЕР. Сравним факторные нагрузки, полученные методом PCA без вращения и с вращением

Матрица компонент[a]

	Компонента		
	1	2	3
аналогии	,668		
счет в уме	,657	-,470	
умозаключения	,657		
числовые ряды	,634		-,420
пропущенные слова	,633		
скрытые фигуры	,615		,471
геометрическое сложение	,589		,456
исключение изображений	,552		
заучивание слов		,732	
осведомленность	,429	,583	
понятливость	,507		,545

Метод выделения: Анализ методом главных компонент.

a. Извлеченных компонент: 3

Матрица повернутых компонент[a]

	Компонента		
	1	2	3
счет в уме	,804		
аналогии	,769		
числовые ряды	,762		
умозаключения	,696		
заучивание слов		,829	
осведомленность		,746	
пропущенные слова		,630	
понятливость			,739
скрытые фигуры			,737
геометрическое сложение	,421		,692
исключение изображений		,487	,506

Метод выделения: Анализ методом главных компонент.

Метод вращения: Варимакс с нормализацией Кайзера.

a. Вращение сошлось за 5 итераций.

Матрица компонент[a]

	Компонента		
	1	2	3
аналогии	0.668		
геометрическое сложение	0.589		0.456
заучивание слов		0.732	
исключение изображений	0.552		
осведомленность	0.429	0.583	
понятливость	0.507		0.545
пропущенные слова	0.633		
скрытые фигуры	0.615		0.471
счет в уме	0.657	-0.47	
умозаключения	0.657		
числовые ряды	0.634		-0.42

Метод выделения: Анализ методом главных компонент.

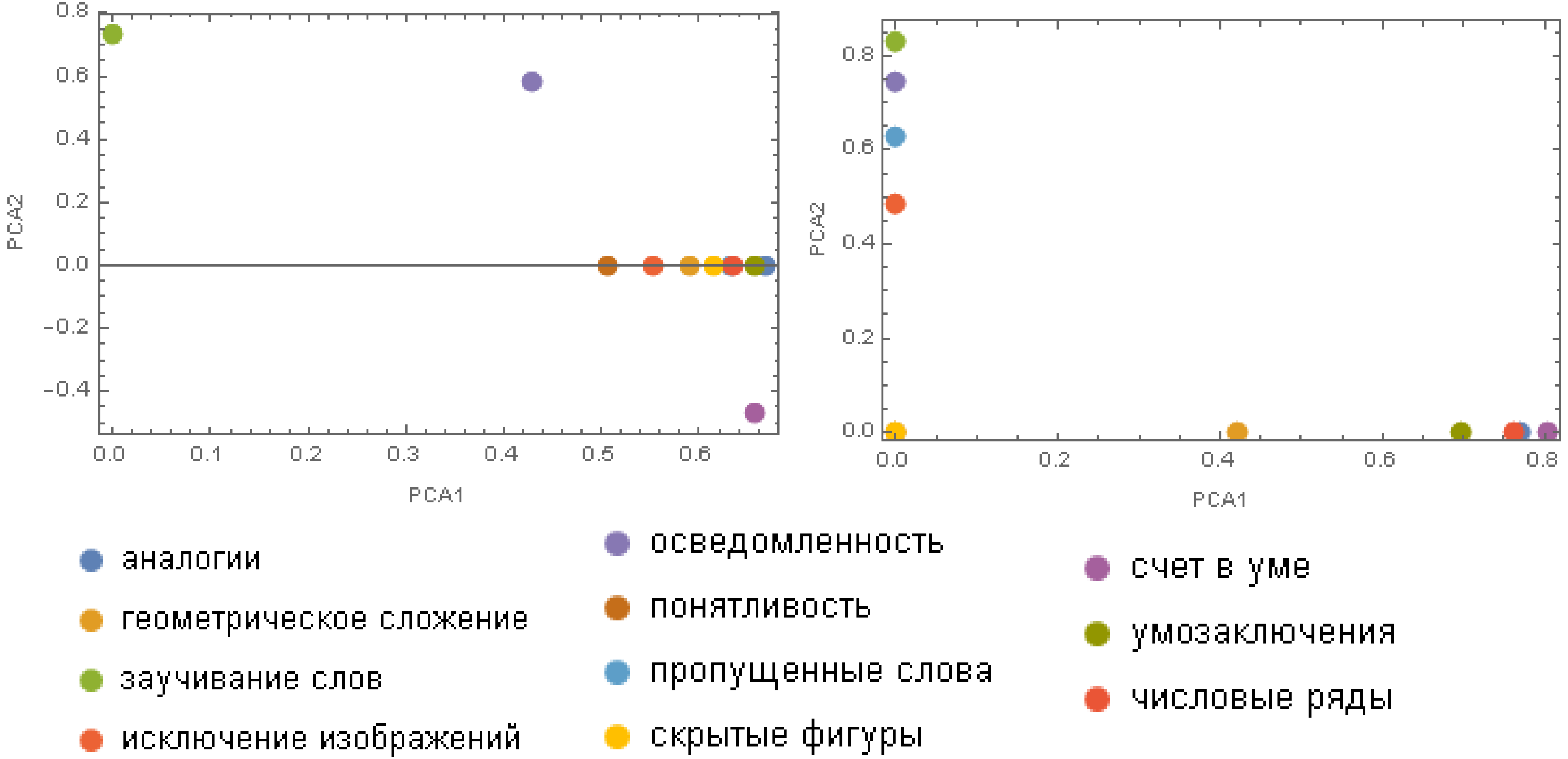
а. Извлеченных компонент: 3

Матрица повернутых компонент[a]

	Компонента		
	1	2	3
аналогии	0.769		
геометрическое сложение	0.421		0.692
заучивание слов		0.829	
исключение изображений		0.487	0.506
осведомленность		0.746	
понятливость			0.739
пропущенные слова		0.63	
скрытые фигуры			0.737
счет в уме	0.804		
умозаключения	0.696		
числовые ряды	0.762		

Метод выделения: Анализ методом главных компонент.

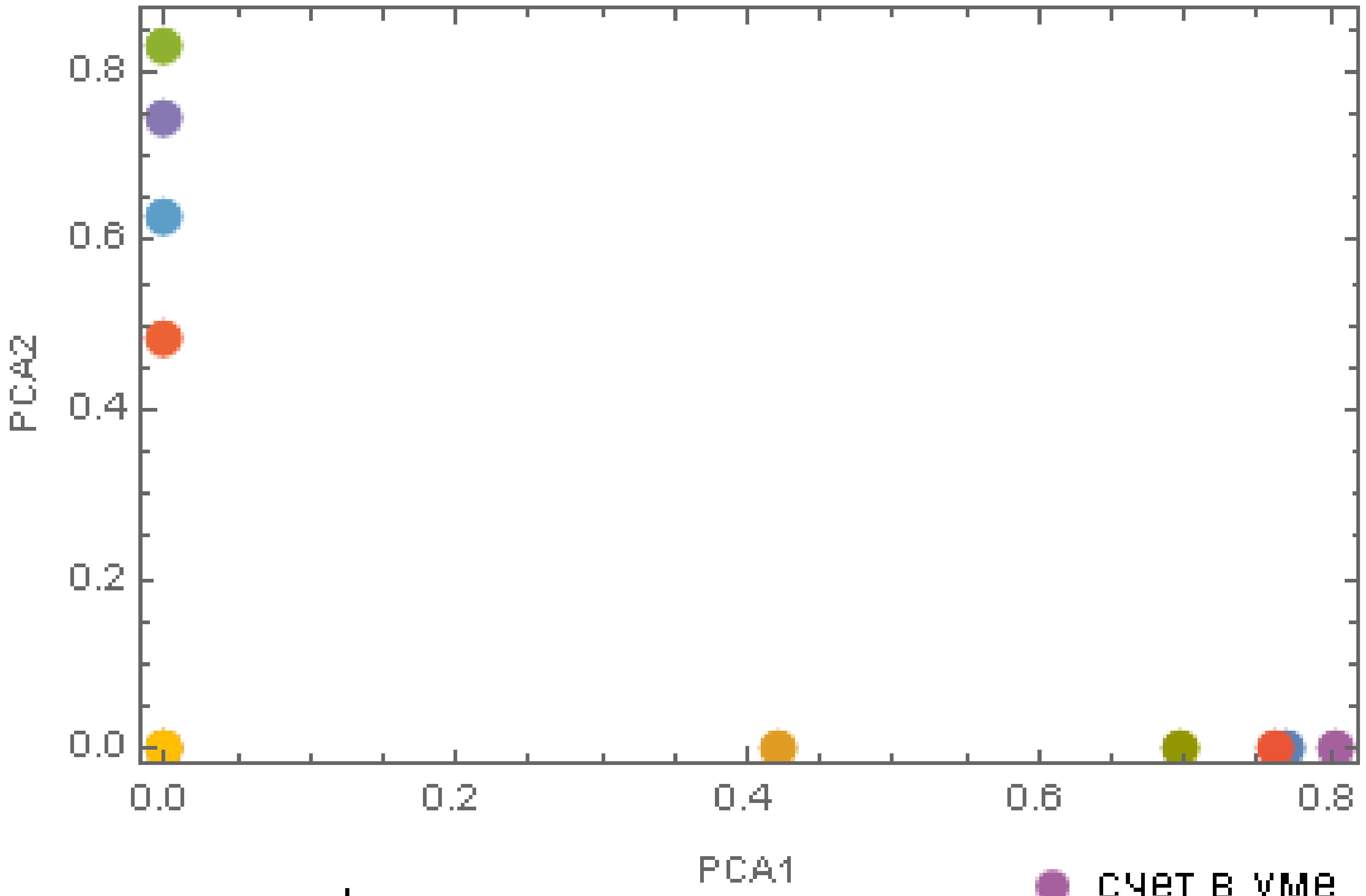
Метод вращения: Варимакс с нормализацией Кайзера.



Переменные в пространстве двух факторов . Получено методом PCA без вращения (слева) и с вращением (справа). В каком случае факторы легче интерпретируются?

- заучивание слов
- исключение изображений
- осведомленность
- пропущенные слова

Вербальные способности,
2-ой фактор



- скрытые фигуры
- понятливость
- геометрическое сложение

Не вербальные способности,
3-ий фактор

- счет в уме
- аналогии
- числовые ряды
- умозаключения

Математико-логические способности
1-ый фактор

ИТОГ:

- Под названием факторный анализ скрывается большое разнообразие методов.
- 1. Проверить, нет ли пропущенных значений.
- Исходные данные для команд – исходная таблица данных, или корреляционная матрица, или ковариационная матрица.
- 2. Выбрать **факторную модель**.
- 3. Выбрать количество компонент/факторов.
- 4. **Выделить компоненты/факторы (факторы, факторные нагрузки),** первичное факторное решение.
- 5. Выполнить **вращение,** вторичное факторное решение.
- 6. **Интерпретировать** результаты.
- 7. Найти значения компонент/факторов.