

Факторный анализ

Описание

Компонент применяется с целью понижения размерности пространства факторов. Это необходимо в случаях, когда входные факторы коррелированы друг с другом, т. е. взаимозависимы. В факторном анализе речь идет о выделении из множества измеряемых характеристик объекта факторов, более емко отражающих свойства объекта.

Первым этапом факторного анализа является выбор новых признаков, которые являются линейными комбинациями прежних и «вбирают» в себя большую часть общей изменчивости входных факторов. Поэтому они содержат большую часть информации, заключенной в первоначальных данных.

В компоненте «Факторный анализ» это осуществляется с помощью метода главных компонент. Он сводится к выбору новой ортогональной системы координат в пространстве наблюдений. В качестве первой главной компоненты избирают направление, вдоль которого массив данных имеет наибольший разброс, а выбор каждой последующей происходит так, чтобы разброс данных вдоль нее был максимальным, и, чтобы она была ортогональна другим главным компонентам, выбранным прежде.

▼ **Пример:**

Проведем факторный анализ небольшой таблицы, содержащей некоторые статистические данные по регионам:

Исходная таблица:

Регион	Население (тыс. чел.)	Доля городского населения, %	Занятых в экономике (тыс. чел.)	Доходы на человека (руб./ мес.)
Тамбовск ая обл.	1269	58,4	532,4	1187,1
Пензенс кая обл.	1531	64,6	674,5	936,8
Ростовск ая обл.	4358	67,6	1811,8	1033,6
Читинск ая обл.	1259	62,4	439,5	472,9

Регион	Население (тыс. чел.)	Доля городского населения, %	Занятых в экономике (тыс. чел.)	Доходы на человека (руб./ мес.)
Чукотски й а. о.	72	67,9	33,8	963,7

Применим компонент «Факторный анализ» к исходной таблице, создав соответствующий узел и задав ему параметр Используемое» для столбцов «Население (тыс. чел.)», «Доля городского населения, %», «Занятых в экономике (тыс. чел.)», «Доходы на человека (руб./мес.)» в настройках входного порта.

В мастере настройки зададим параметры: «Критерий значимости фактора» — «По собственному значению» с порогом собственного значения 1, «Метод вращения» — «Без вращения». Ограничим число выходных факторов равным 2.

Таблица факторов:

Фак тор 1	Фак тор 2	Регион	Населен ие (тыс. чел.)	Доля городского населения, %	Занятых в экономике (тыс. чел.)	Доходы на человека (руб./мес.)
-0,4 0	-1,8 7	Тамбов ская обл.	1269	58,4	532,4	1187,1
-0,0 4	0,02	Пензен ская обл.	1531	64,6	674,5	936,8
1,93	0,23	Ростов ская обл.	4358	67,6	1811,8	1033,6
-0,7 1	1,06	Читинс кая обл.	1259	62,4	439,5	472,9
-0,7 8	0,55	Чукотс кий а. о.	72	67,9	33,8	963,7

Таблица факторных нагрузок:

Метка	Фактор1	Фактор2
Население (тыс. чел.)	0,97	0,01
Доля городского населения, %	0,41	0,67

Метка	Фактор1	Фактор2
Занятых в экономике (тыс. чел.)	0,98	-0,03
Доходы на человека (руб./мес.)	0,33	-0,77

Таким образом, опираясь на существующую взаимосвязь факторов (корреляцию), мы понизили размерность пространства факторов исходной таблицы с 4 до 2.

Порты

Вход

- ☐ **Входной источник данных** (таблица данных). В настройках этого порта следует выставить назначение  «Используемое» для полей, данные которых следует подвергнуть факторному анализу.

Выход

- ☐ **Факторы** — содержит исходную таблицу, к которой добавлены поля факторов.
- ☐ **Факторные нагрузки** — содержит коэффициенты корреляции между исходными признаками(используемыми полями) и выделенными в ходе анализа факторами, показывающие степень влияния каждого признака на новый фактор.

Мастер настройки

- Критерий значимости факторов:
 - ☐ **По собственному значению** — отбираются только факторы с собственными значениями равными или большими 1. Считается, что те факторы, у которых этот показатель меньше 1, не вносят значительного вклада в объяснение результата.
 - ☐ **По дисперсии** — факторы отбираются по доле объясняемой несмещённой дисперсии. В этом случае выбирают столько факторов, чтобы в сумме они объясняли не менее 70-75% дисперсии. В отдельных вариантах порог несмещенной дисперсии может достигать 85-90%.
 - ☐ **Задать число факторов** — количество значимых факторов выбирается аналитиком самостоятельно.

Значение каждого критерия можно задать в специальном поле справа от соответствующей ему радиокнопки.

- Метод получения окончательного решения
 - ☐ **Без вращения** — исходные факторы, полученные методом главных компонент, остаются без изменений.

- **Варимакс** — критерием является упрощение описания каждого фактора. В результате максимизируется нагрузка на каждый фактор относительно небольшого числа переменных, а факторные нагрузки остальных переменных минимизируются. Рекомендуется выбирать, когда требуется обеспечить высокую интерпретируемость результатов факторного анализа.
- **Квартимакс** — данный критерий упрощает описание каждой переменной, то есть уменьшает число факторов, связанных с этой переменной.
- **Ограничить число выходных факторов** — необязательный флаг, при активации которого можно задать ограничивающее число выходных факторов в соответствующем поле области.