



Факторный анализ



Факторный анализ: понятие и назначение

Основные этапы факторного анализа

- 2.1. Вычисление корреляционной матрицы
- 2.2. Извлечение факторов
- 2.3. Выбор и вращение факторов
- 2.4. Интерпретация факторов
- 2.5. Вычисление значений факторов
- 2.6. Оценка качества модели



Факторный анализ: понятие и назначение

Факторный анализ - это процедура, с помощью которой большое число переменных, относящихся к имеющимся наблюдениям, сводят к меньшему количеству независимых влияющих величин, называемых факторами: в один фактор объединяются переменные, сильно коррелирующие между собой.

переменные из разных факторов слабо коррелируют между собой.

Факторный анализ классифицирует признаки (переменные), описывающие наблюдения.

Фактор (Factor) - латентная (скрытая) переменная, конструируемая таким образом, чтобы можно было объяснить корреляцию между набором имеющихся переменных.

Концепция факторного анализа заключается в «сжатии» информации.



Факторный анализ: понятие и назначение

Цель факторного анализа – сокращение числа переменных на основе их классификации и определения структуры взаимосвязей между ними.

Благодаря сокращению числа переменных вместо исходного набора переменных появляется возможность анализировать данные по выделенным факторам, число которых значительно меньше исходного числа взаимосвязанных переменных.

Социологический смысл анализа – измеряемые эмпирические показатели считаются следствием других, глубинных, скрытых от непосредственного измерения характеристик (латентных переменных)..



Факторный анализ: понятие и назначение

Пример применения факторного анализа

Расчет прибыли от продаж компании. Прибыль зависит от четырех основных факторов:

- объема продаж
 - ассортимента реализованной продукции
 - себестоимости продукции
 - цены реализации продукции.
- С помощью факторного анализа можно рассчитать, как каждый из перечисленных факторов влияет на величину прибыли компании, и, исходя из полученных результатов, выработать пути максимизации прибыли.

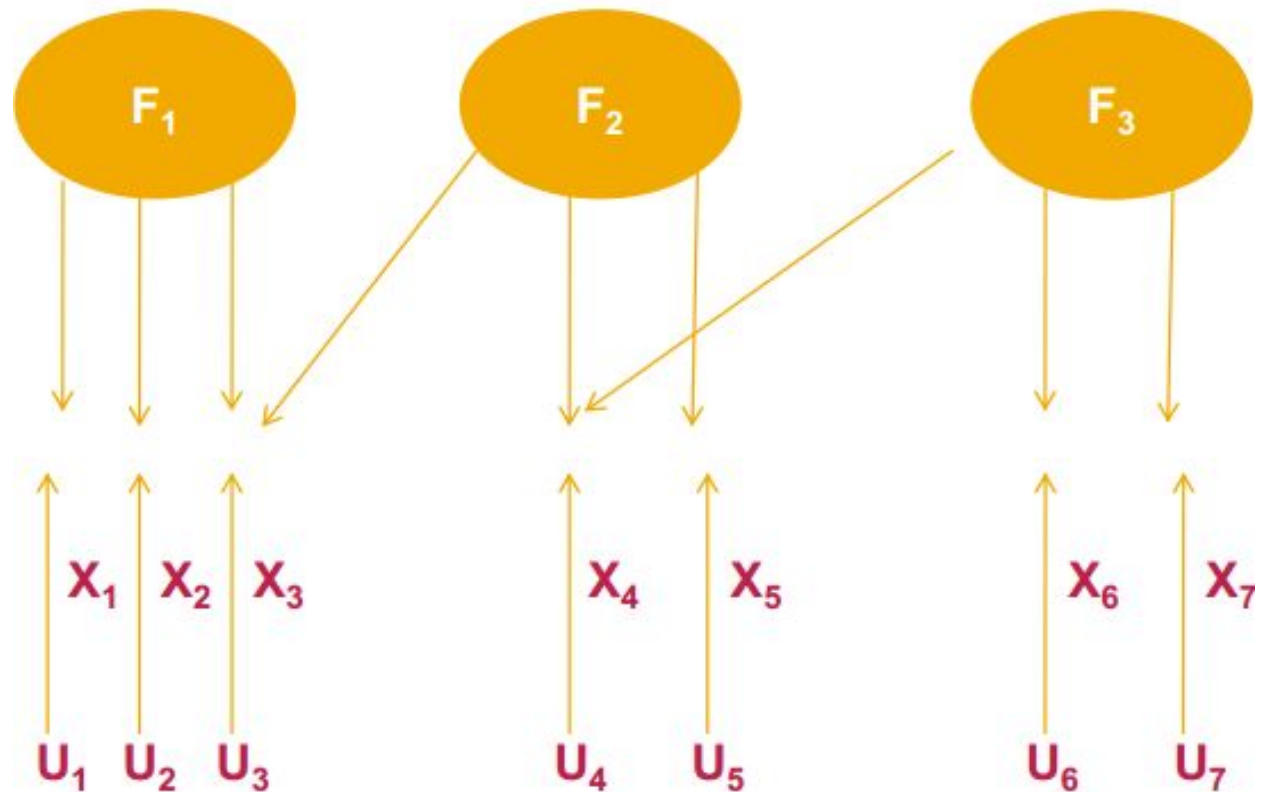
Аналогичным образом можно рассчитать и проанализировать затраты на производство продукции.



Схема факторного анализа

- F** – общие факторы, каждый из которых влияет на определенную совокупность переменных
- X** - переменные, фиксируемые на основании ответов
- U** - уникальные факторы, каждый из которых влияет только на одну переменную

Схема факторного анализа



Факторный анализ: понятие и назначение

Все признаки должны быть количественными переменными (интервальными либо метрическими).

Число наблюдений должно быть минимум в два раза больше числа переменных.

Выборка должна быть однородна.

Исходные переменные должны быть распределены симметрично.

Номинальные переменные должны быть переведены в дихотомические (переменные, имеющие только две категории).



Основные этапы факторного анализа

Порядок выполнения факторного анализа

Процедура факторного анализа состоит из четырех основных стадий:

1. Вычисление корреляционной матрицы для всех переменных, участвующих в анализе.
2. Извлечение факторов.
3. Выбор факторов и вращение факторов для создания упрощенной структуры.
4. Интерпретация факторов.



Вычисление корреляционной матрицы

- Первая операция, которая производится при выполнении факторного анализа — вычисление корреляционной матрицы для переменных, участвующих в анализе.
- Для проведения факторного анализа совсем не обязательно специально строить корреляционную матрицу: при необходимости программа создаст ее сама на основе данных файла.
- иногда не нужны даже исходные данные — достаточно иметь корреляционную матрицу, которая в этом случае вводится в командный файл



Проведение анализа

«Анализ» → «Многомерный разведочный анализ» →
«Факторный анализ»

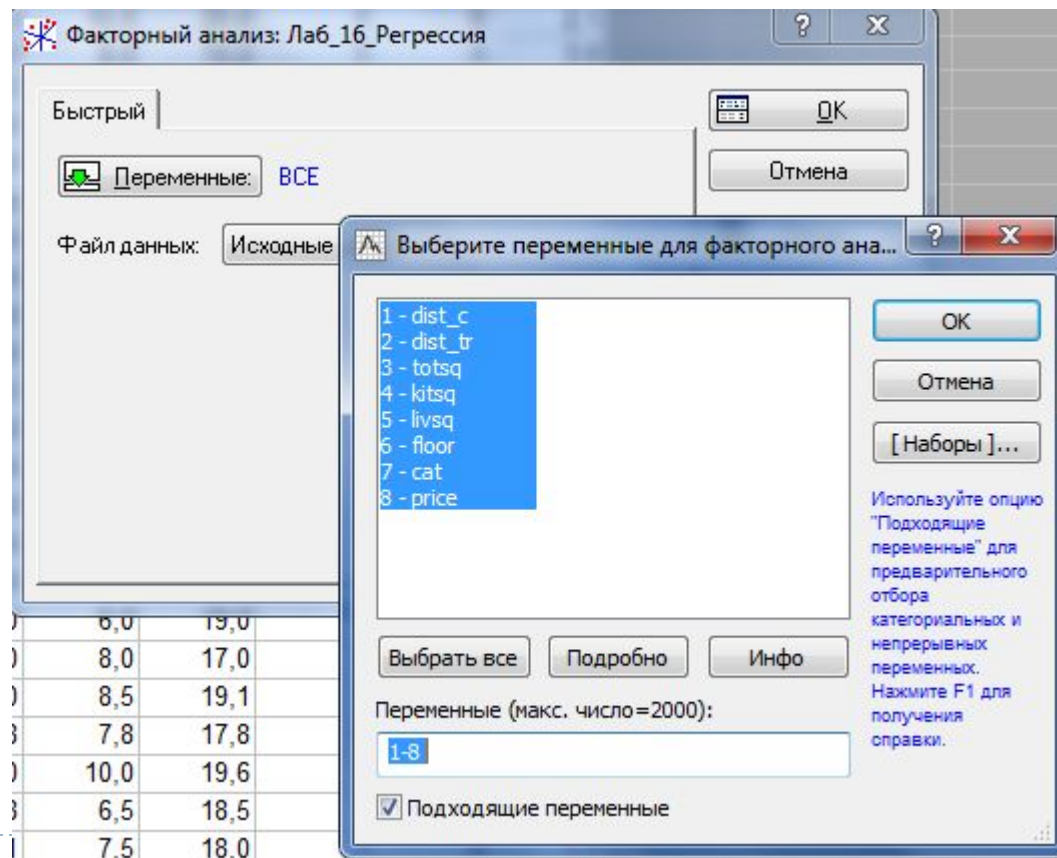
The screenshot shows the SPSS interface with a data table and a 'Factor Analysis' dialog box. The data table has the following columns: 1 dist_c, 2 dist_tr, 3 totsq, 4 kitsq, 5 livsq, 6 floor, 7 cat, 8 price. The dialog box is titled 'Факторный анализ: Лаб_16_Регрессия' and has the following settings:

- Быстрый: (unchecked)
- Переменные: ВСЕ
- Файл данных: Исходные данные
- Опции: (unchecked)
- Данные: (unchecked)
- Удаление ПД: Построчное, Парное, Замена средним

	1 dist_c	2 dist_tr	3 totsq	4 kitsq	5 livsq	6 floor	7 cat	8 price
1,000000	3,6	10	35,0					
2,000000	3,6	10	37,0					
3,000000	3,6	15	32,0					
4,000000	3,6	15	35,0					
5,000000	3,6	5	43,0					
6,000000	3,6	10	30,0					
7,000000	3,6	10	30,0					
8,000000	3,6	15	32,0					
9,000000	9,8	2	38,0					
10,000000	9,8	3	33,0					
11,000000	9,8	5	37,0					
12,000000	9,8	10	33,0					
13,000000	9,8	3	36,0					
14,000000	9,8	5	33,0					
15,000000	9,8	1	39,0					
16,000000	9,8	5	38,0					
17,000000	9,8	7	32,0					

Проведение анализа

«Анализ» → «Многомерный разведочный анализ» →
«Факторный анализ» → Переменные → Выбрать все



Проведение анализа

«Анализ» —> «Многомерный разведочный анализ» —>

«Факторный анализ» —> Переменные —> Выбрать все —> ОК

The screenshot shows the STATISTICA software interface. The main window displays a data table with the following columns: 1 dist_c, 2 dist_tr, 3 totsq, 4 kitsq, 5 livsq, 6 floor, 7 cat, 8 price. The data rows range from 1,000,000 to 22,000,000. A dialog box titled "Задайте метод выделения факторов: Лаб_16_Регрессия" is open, showing the "Быстрый" (Quick) tab. The dialog box contains the following text: "ПД построчно удалены", "66 наблюдений выбрано", "66 набл. принято к обработке", "Матрица корреляций вычислена для 8 перем.". The "Максимальное число факторов:" (Maximum number of factors) is set to 3, and the "Минимальное соб. знач." (Minimum eigenvalue) is set to 1,000. A red circle highlights the "Максимальное число факторов:" field. A text box at the bottom of the dialog box contains the text: "Следует задать максимальное число факторов (2 или 3)".

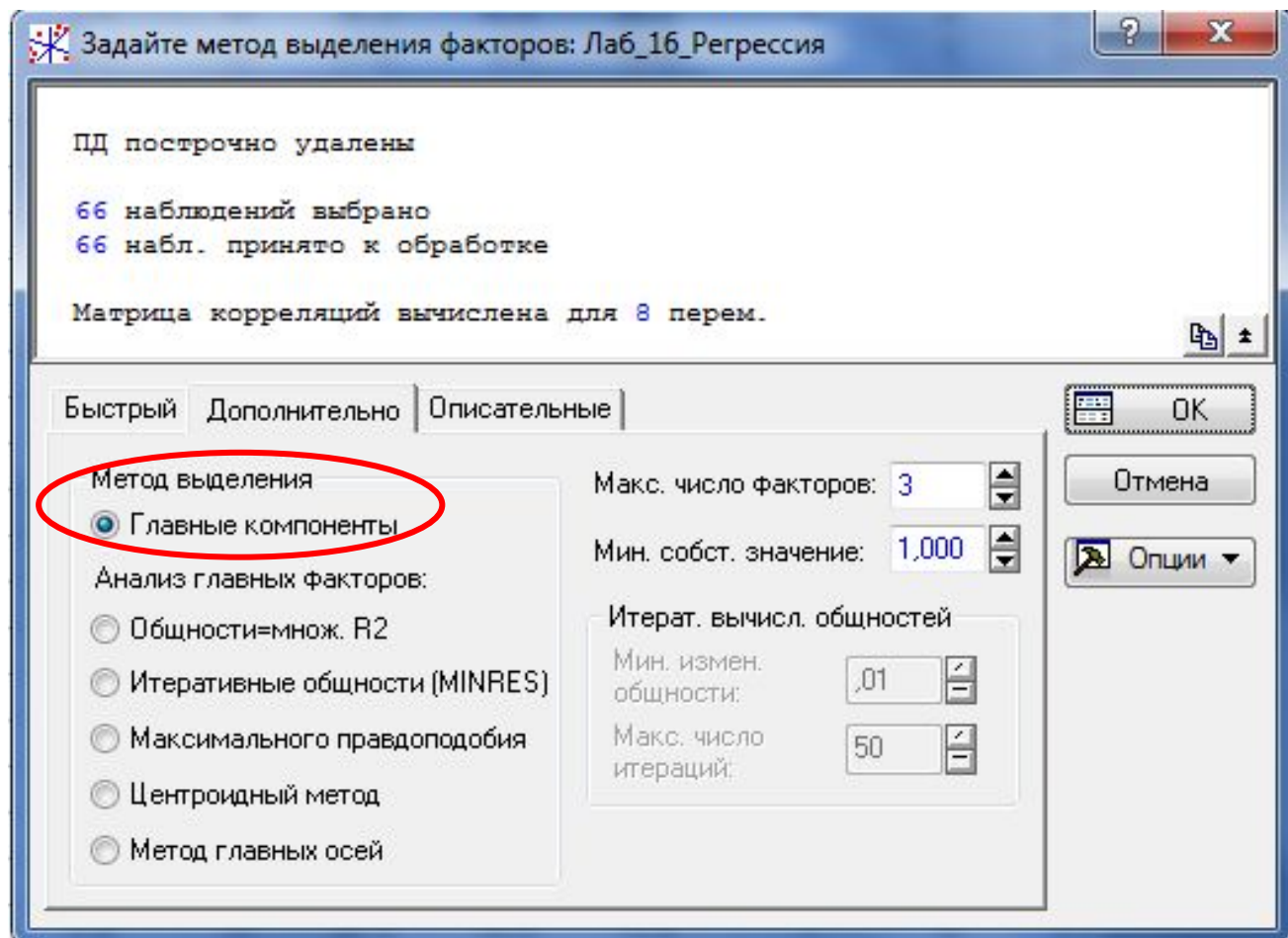
	1	2	3	4	5	6	7	8
	dist_c	dist_tr	totsq	kitsq	livsq	floor	cat	price
1,000000	3,6	10	35,3	10,0	20,0	1	0	2,1
2,000000	3,6	10	37,0	8,0	19,0	1	1	3,0
3,000000	3,6	15	32,2					
4,000000	3,6	15	35,0					
5,000000	3,6	5	43,0					
6,000000	3,6	10	30,0					
7,000000	3,6	10	30,0					
8,000000	3,6	15	32,0					
9,000000	9,8	2	38,0					
10,000000	9,8	3	33,0					
11,000000	9,8	5	37,5					
12,000000	9,8	10	33,0					
13,000000	9,8	3	36,0					
14,000000	9,8	5	33,0					
15,000000	9,8	1	39,0					
16,000000	9,8	5	38,0					
17,000000	9,8	7	33,0					
18,000000	9,8	10	32,0					
19,000000	9,8	10	38,0					
20,000000	9,8	7	34,8					
21,000000	9,8	15	35,0					
22,000000	9,8	3	32,8					
23,000000	9,8	5	32,1					

Открывается
окно на
вкладке
Быстрый

Следует задать
максимальное число
факторов (2 или 3)

Выбор метода выделения факторов

Вкладка Дополнительно → Главные компоненты → ОК



Методы факторного анализа

- Наиболее популярным методом является **метод главных компонент** (Principal Components Analysis) основан на определении минимального числа факторов, вносящих наибольший вклад в дисперсию данных.
- Они называются **главными компонентами**.

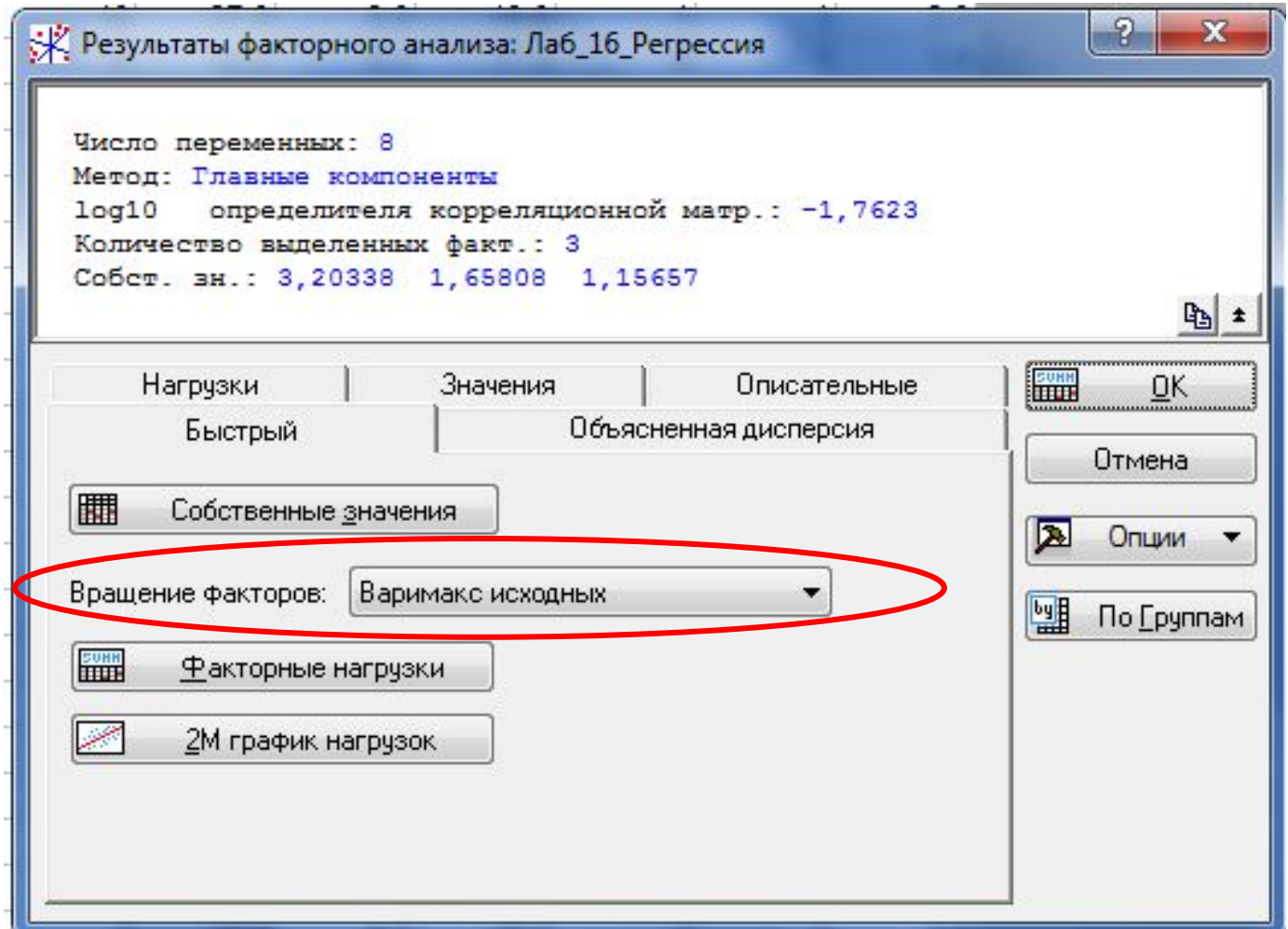
The screenshot shows the STATISTICA software interface. The main window displays a regression analysis for 'Лаб_16 Регрессия'. The data table includes variables: dist_c, dist_tr, totsq, kitsq, livsq, floor, cat, and price. A dialog box titled 'Результаты факторного анализа: Лаб_16 Регрессия' is open, showing the following results:

- Число переменных: 8
- Метод: Главные компоненты
- log10 определителя корреляционной matr.: -1,7623
- Количество выделенных факт.: 3
- Собст. зн.: 3,20338 1,65808 1,15657

The dialog box also shows options for 'Нагрузки' (Loadings) and 'Значения' (Values), with 'Быстрый' (Fast) selected. The 'Объясненная дисперсия' (Explained variance) is shown as 3,20338, 1,65808, and 1,15657. The 'Вращение факторов' (Factor rotation) is set to 'Без вращения' (No rotation). Buttons for 'Собственные значения' (Eigenvalues), 'Факторные нагрузки' (Factor loadings), and '2M график нагрузок' (2M loading plot) are visible.

Вращение факторов

- Наиболее популярным методом вращения факторов является **Варимакс**



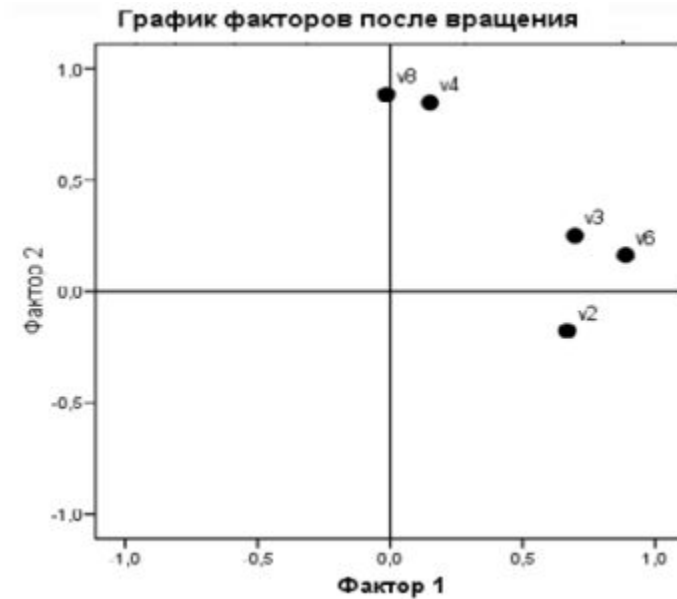
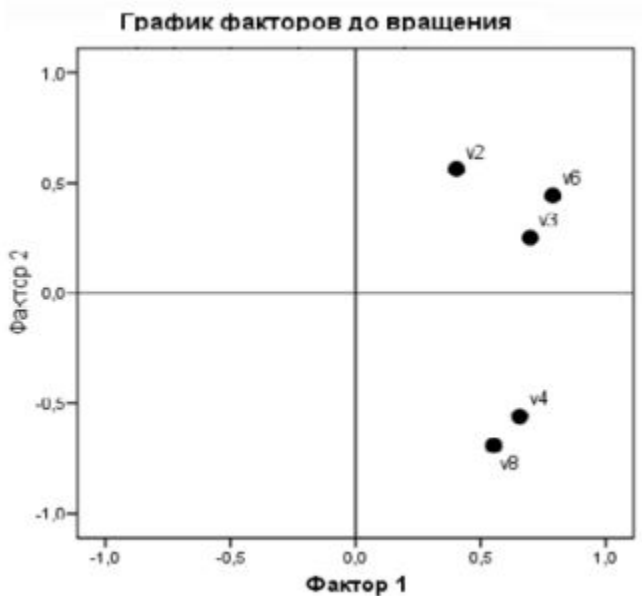
Вращение факторов

- Целью вращения является получение простой структуры, которой соответствует большое значение нагрузки каждой переменной только по одному фактору и малое по всем остальным факторам.
- Нагрузка отражает связь между переменной и фактором, являясь подобием коэффициента корреляции.
- Значение нагрузки лежит в пределах $(-1; 1)$.



Вращение факторов

- До вращения точки, соответствующие переменным, расположены на удалении от осей факторов.
- После поворота осей переменные оказываются вблизи осей, что соответствует максимальной нагрузке каждой переменной только по одному фактору.



Извлечение факторов

- Начинается с подсчета суммарного разброса значений всех участвующих в анализе переменных. Для этого «суммарного разброса» непросто подобрать логическую интерпретацию, однако он является вполне строго определенной математической величиной.
- Извлечение факторов заключается в выборе взаимодействующих переменных, чья взаимная корреляция обуславливает наибольшую долю общей дисперсии. Эти переменные образуют первый фактор.
- Затем первый фактор исключается и из оставшегося множества переменных снова выбираются те, чье взаимодействие определяет наибольшую долю оставшейся общей дисперсии. Эти переменные образуют второй фактор.
- Процедура извлечения факторов продолжается до тех пор, пока не будет исчерпана вся общая дисперсия переменных.
- После того как процедура извлекает каждый фактор, напротив его номера появляется собственное значение, которое пропорционально доле общей дисперсии, определяемой данным фактором.



Извлечение факторов

Значение общности равно доле дисперсии переменной, обусловленной совокупным влиянием факторов:

- Если равно 0 – фактор не влияет на переменную.
- Если равно 1 – дисперсия переменной целиком определяется выделяемым фактором

STATISTICA - [Рабочая книга2* - Фактор.нагрузки (Вар.исходн.) (Лаб_16_Регрессия)]

Файл Правка Вид Вставка Формат Анализ Добыча Данных Графика Сервис

Добавить в Рабочую книгу

Arial 10 B I U

Рабочая книга2*
Факторный анализ
Результаты
Фактор

Перемен.	Фактор 1	Фактор 2	Фактор 3
dist_c	-0,092075	-0,869899	0,144777
dist_tr	0,032142	0,123675	-0,794599
totsq	0,931282	0,170963	-0,067693
kitsq	0,820329	-0,237390	-0,027127
livsq	0,680877	0,356475	0,135105
floor	0,042582	-0,073702	0,819716
cat	0,158087	0,880174	-0,069213
price	0,770782	0,505216	0,105267
Общ.дис.	2,634241	2,020059	1,363726
Доля общ	0,329280	0,252507	0,170466

Интерпретация факторов

- Возможно графическое изображение переменных, координатами которых служат величины факторных нагрузок. Так, в конце оси расположены переменные, которые имеют большие нагрузки только в связи с этим фактором и, следовательно, характеризуют его.
 - Переменные в начале координат имеют небольшие нагрузки в связи с обоими факторами.
 - Переменные, расположенные вдали от осей, связаны с обоими факторами.
 - Если фактор нельзя четко определить с точки зрения связи с исходными переменными, то его следует пометить как неопределяемый или генеральный (общий для всех переменных).
-



Интерпретация факторов

- Красным выделены нагрузки более 0,70
- Интерпретировать следует связь переменных с выделенной нагрузкой во всех трёх факторах

Рабочая книга2* - Фактор.нагрузки (Вар.исходн.) (Лаб_16_Регрессия)

Фактор.нагрузки (Вар.исходн.) (Лаб_16_Регрессия)
Выделение: Главные компоненты
(Отмечены нагрузки >,700000)

Перемен.	Фактор 1	Фактор 2	Фактор 3
dist_c	-0,092075	-0,869899	0,144777
dist_tr	0,032142	0,123675	-0,794599
totsq	0,931282	0,170963	-0,067693
kitsq	0,820329	-0,237390	-0,027127
livsq	0,680877	0,356475	0,135105
floor	0,042582	-0,073702	0,819716
cat	0,158087	0,880174	-0,069212
price	0,770782	0,505216	0,105267
Общ. дис.	2,634241	2,020059	1,363726
Доля общ	0,329280	0,252507	0,170466

STATISTICA - [Данные: Автомобили_Факторный анализ.sta* (10v * 23с)]

Файл Правка Вид Вставка Формат Анализ Добыча Данных Графика Сервис Данные Окно Справка

Продолжить... Ctrl+R

Основные статистики и таблицы
 Множественная регрессия
 Дисперсионный анализ (ДА)
 Непараметрическая статистика
 Подгонка распределений
 Подгонка и моделирование

Углубленные методы анализа
Многомерный разведочный анализ
 Промышленная статистика и Шесть Сигма
 Анализ мощности
 Автоматизированные Нейронные Сети
 PLS, PCA, Многомерный/Пакетный SPC
 Оценка дисперсии и точности

Блочные статистики
 STATISTICA Visual Basic
 Пакетный анализ

Вероятностный калькулятор

Класстерный анализ
Факторный анализ
 Анализ главных компонент и классификация
 Канонический анализ
 Надежность и позиционный анализ
 Деревья классификации
 Анализ соответствий
 Многомерное шкалирование
 Дискриминантный анализ
 Общие модели дискриминантного анализа

	4 щность л.с.	5 максимальная скорость	6 разгон до 100 км/ч	7 в городе л/100 км	8 на шоссе л/100 км	9 средн. л/100 км	10 полная масса			
Mercedes-Benz S-класс	204	244	6,4	3,9	3,4	3,6	2195			
BMW 7-серия	265	250	5,8	5,7	4,3	4,8	2540			
Opel Insignia 2.0				6,1	4,2	4,6	2063			
Maserati Quattroporte S				7,8	5,2	6,2	1885			
Audi A8				6,4	6,2	6,3	2520			
VW Passat 1.8				7,1	5	5,8	2030			
Porsche Panamera				9,8	6,3	7,6	2390			
Mazda 6 2.0				6,6	4,2	5,1	1625			
Cadillac ATS				12	6,3	8,4	2430			
Alfa Romeo 159				11,8	6	8,1	1980			
Citroen C5 1.6				11,1	5,8	7,7	2026			
Jaguar XJ				13,6	6,8	9,3	2265			
Ford Mondeo 2.0				11,6	6	8	1915			
Volvo C70				230	235	8	14	6,7	9,4	2100
Lexus LS 460	5030	1465	4608	380	250	5,7	12,3	8,7	11,1	2400
Kia Quoris	5090	1490	3778	290	240	7,3	14,7	7,8	10,3	2430
Aston Martin DB11	4739	1279	5200	608	322	3,9	15	8,1	11,4	1770
VW Phaeton	5055	1450	2967	233	234	8,9	12,8	8,1	9,7	2252
Renault Latitude 2.5	4897	1483	2495	177	209	10,7	13,4	7,7	9,7	2110
Rolls-Royce Phantom	6092	1640	6749	460	240	6,1	23	10,2	14,9	3170
Lamborghini Gallardo	4345	1184	5240	560	324	4	20,8	9,7	13,8	1550
Saab 9-5	4922	1791	4157	279	191	8	15	11	13	2608
Bentley Continental	5290	1398	5998	560	312	5,2	26,2	11,9	17,1	2940

Начать Факторный анализ

Автомобили_ Н14.П7 14 Выбор:OFF Вес:OFF CAP NUM REC