



МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ГЕОГРАФИИ


Карпиченко Александр
Александрович

*доцент кафедры почвоведения и
геоинформационных систем*





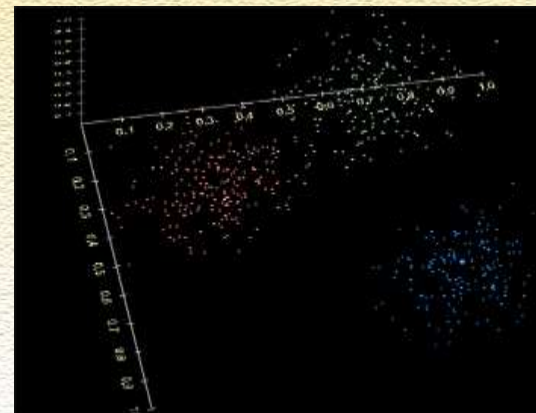
Литература

- elib.bsu.by
 - Математические методы в географии: учебно-методическое пособие / Н. К. Чертко, А. А. Карпиченко. – Минск: БГУ, 2009.
- 

7. ФАКТОРНЫЙ АНАЛИЗ

При изучении взаимного влияния многих процессов и явлений в последнее время все чаще обращаются к методам многомерного статистического анализа, в частности факторного анализа. Методы многомерного статистического анализа практически всегда выполняются на специализированном ПО.

Факторный анализ основывается на использовании статистических знаний (вычислении стандартных отклонений, знании корреляционного и регрессионного анализов). В большинстве случаев исследуется система корреляций, отраженных в корреляционной матрице.



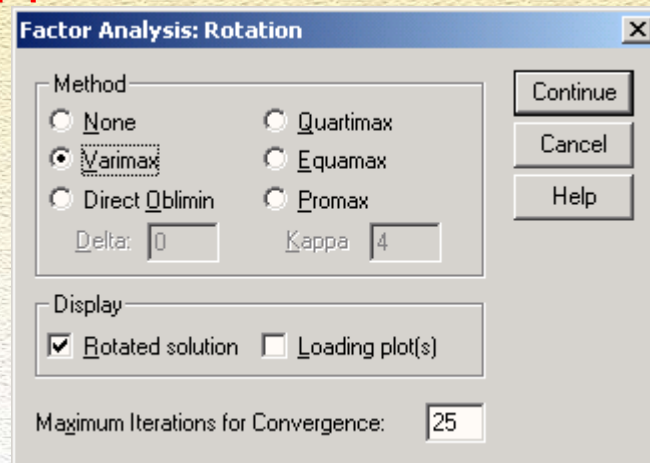
7. ФАКТОРНЫЙ АНАЛИЗ

Факторный анализ представляет собой ветвь математической статистики, цель которого – разработка моделей, понятий и методов, позволяющих анализировать и интерпретировать массивы экспериментальных данных независимо от их физической природы. Анализ данных включает краткое описание распределения объектов, установление взаимоотношения процессов и явлений, отражающихся в виде **параметров**.

Value	Eigenvalues (factor) Extraction: Principal components			
	Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative %
1	2,893561	36,16951	2,893561	36,16951
2	1,708925	21,36156	4,602486	57,53108

7. ФАКТОРНЫЙ АНАЛИЗ

Используемый набор моделей и методов предназначен для «сжатия» информации, содержащейся в корреляционной матрице. В основе различных моделей факторного анализа лежит следующая гипотеза: параметры – это косвенные характеристики объекта или явления и представляют в совокупности тот или иной фактор. В связи с этим задача факторного анализа состоит в том, чтобы показать наблюдаемые параметры в виде линейных комбинаций факторов.



7. ФАКТОРНЫЙ АНАЛИЗ


Изменение фактора не всегда одинаково отражается на параметрах, поэтому среди последних могут быть выделены группы, реагирующие на каждый из факторов порознь. **Параметры, входящие в одну и ту же группу, сильно коррелируют между собой; параметры, входящие в разные группы, слабо коррелируют между собой.** Задача выявления факторов понимается как разбиение параметров на группы таким образом, чтобы можно было описать взаимоотношения между параметрами.

Variable	Factor Loadings (Varimax raw) (factor)	
	Factor 1	Factor 2
органические удобрения, т/га	0,872127	-0,067592
минеральные удобрения, ц/га	0,718471	-0,441738
дозы извести, т/га	0,972194	-0,040124
пестициды, кг/га	-0,115877	0,040150
гумус, т/га	-0,022341	0,820951
гидролитическая кислотность (Н), мэкв/100г	-0,138428	0,906199
влажность почвы, %	0,075795	-0,565191
физическая глина, %	0,552192	0,138743
Expl.Var	2,566717	2,036770
Prp.Totl	0,320715	0,254596



7. ФАКТОРНЫЙ АНАЛИЗ

Разработано несколько вариантов факторного анализа с использованием коэффициентов только линейной корреляции. Наиболее употребительны при этом *метод главных компонент*, *метод главных факторов* и *центридный метод*. Определение главных компонент и главных факторов производится с помощью специализированного программного обеспечения.



7. ФАКТОРНЫЙ АНАЛИЗ

Метод факторного отображения используется при решении ряда географических задач: для целей инженерно-географического районирования и количественной оценки влияния природных условий на производство; для организации отдыха и т. д. По матрицам значений факторов можно составлять картосхемы, на основе которых осуществляется территориальный анализ выражения важнейших факторов.

7. ФАКТОРНЫЙ АНАЛИЗ

Наиболее типичной формой представления данных при факторном анализе является *матрица*. Это прямоугольная (или квадратная) таблица чисел, вертикальный ряд которой (столбец) обозначается индексом j , горизонтальный (строка) – индексом i . Любой элемент матрицы обозначается символом a с индексами, первый указывает номер строки, второй – номер столбца, которым соответствует данный элемент (в общем виде a_{ij}). Матрица обозначается прописной буквой (A , B и т. д.). О матрице, имеющей m строк и n столбцов, говорят, что ее порядок составляет $m \cdot n$. Квадратная матрица $n \cdot n$ имеет порядок n . В общем виде матрица записывается следующим образом:

$$A = \begin{vmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{vmatrix}$$



7. ФАКТОРНЫЙ АНАЛИЗ

В факторном анализе с использованием правил матричной алгебры часто встречается операция умножения матриц. Для того чтобы умножить матрицу A на матрицу B , необходимо следующее условие: матрица A должна иметь столько столбцов, сколько строк в матрице B . Сам процесс умножения протекает по правилу «строка на столбец». Это правило означает, что каждый элемент матрицы произведения представляет собой сумму произведений элементов строки первой матрицы на соответствующие элементы столбца второй матрицы. Матрица-произведение будет иметь всегда столько строк, сколько их было в первой матрице, и столько столбцов, сколько их было во второй матрице: $(p \cdot q) \cdot (q \cdot r) = (p \cdot r)$.

$$A = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \times B = \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{vmatrix} = C = \begin{vmatrix} (a_{11}b_{11} + a_{12}b_{21}) & (a_{11}b_{12} + a_{12}b_{22}) & (a_{11}b_{13} + a_{12}b_{23}) \\ (a_{21}b_{11} + a_{22}b_{21}) & (a_{21}b_{12} + a_{22}b_{22}) & (a_{21}b_{13} + a_{22}b_{23}) \\ (a_{31}b_{11} + a_{32}b_{21}) & (a_{31}b_{12} + a_{32}b_{22}) & (a_{31}b_{13} + a_{32}b_{23}) \end{vmatrix}$$


7. ФАКТОРНЫЙ АНАЛИЗ

Элементами исходной матрицы в факторном анализе являются коэффициенты корреляции. В ходе анализа вычисляется также общая дисперсия σ^2 , указывающая, в каких границах находятся значения параметров, которые характеризуют фактор. Кроме общей дисперсии в анализе учитывается факторная дисперсия (общность) и специфическая дисперсия, связанная с некоторой переменной и характеризующая только ее. Дисперсию, обусловленную ошибкой, стремятся свести к минимуму.



7. ФАКТОРНЫЙ АНАЛИЗ

Элементы столбцов факторной матрицы представляют собой **факторные нагрузки**, или **коэффициенты факторного отображения**, выраженные коэффициентами корреляции данной переменной с данным фактором. Таким образом, коэффициенты факторного отображения характеризуют фактор и его влияние на все параметры.

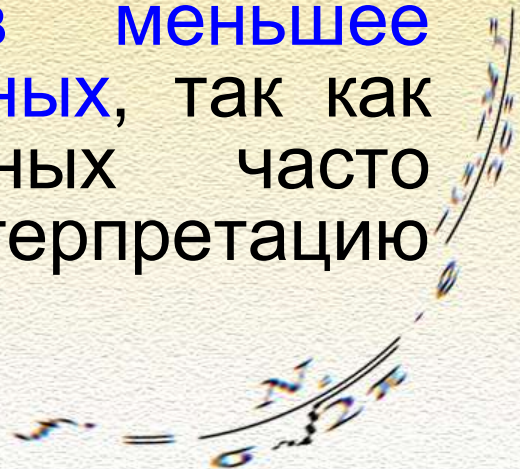




7. ФАКТОРНЫЙ АНАЛИЗ

Часто целями факторного анализа являются:
сокращение числа переменных и
определение структуры взаимосвязей
между переменными. Поэтому факторный
анализ используется или как метод сокращения
данных или как метод классификации.

Метод главных компонент — это метод,
который переводит большое количество
связанных между собой (зависимых,
коррелирующих) переменных в меньшее
количество независимых переменных, так как
большое количество переменных часто
затрудняет анализ и интерпретацию
информации.



7. ФАКТОРНЫЙ АНАЛИЗ

Особенностью метода главных компонент является, во-первых, то, что в ходе вычислительных процедур одновременно получают все главные компоненты и их число первоначально равно числу исходных переменных; во-вторых, постулируется возможность полного разложения дисперсии всех исходных переменных, т.е. ее полное объяснение через латентные факторы (обобщенные признаки).

Value	Eigenvalues (factor) Extraction: Principal components			
	Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative %
1	2,893561	36,16951	2,893561	36,16951
2	1,708925	21,36156	4,602486	57,53108

7. ФАКТОРНЫЙ АНАЛИЗ

Пример. На основе выборки по 395 ландшафтам в пределах водораздельного пространства была получена исходная информация о шести параметрах агроландшафтов. Они включают: 1) органические удобрения; 2) минеральные удобрения; 3) известь; 4) пестициды; 5) содержание гумуса в пахотном горизонте; 6) кислотность почв. Следует определить, какова роль этих параметров в эволюции агроландшафтов. Вполне возможно, что данные показатели будут коррелировать между собой.

Если переменных в исследовании слишком много (x_1, x_2, \dots, x_p), а некоторые из них взаимосвязаны, то у исследователя иногда возникает желание уменьшить сложность данных, сократив количество переменных.

7. ФАКТОРНЫЙ АНАЛИЗ

Для этого и служит **метод главных компонент**, который создает несколько новых переменных y_1, y_2, \dots, y_p , каждая из которых является линейной комбинацией первоначальных переменных x_1, x_2, \dots, x_p :

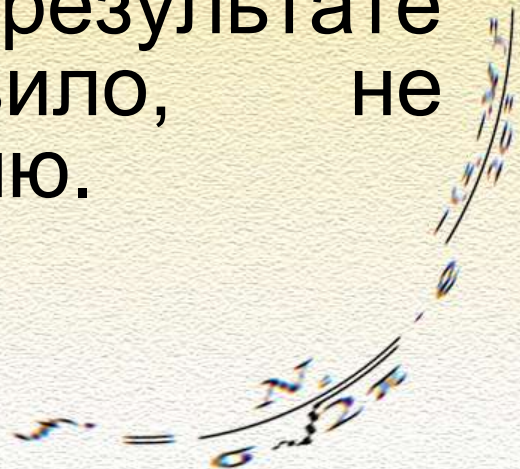
$$\begin{aligned}y_1 &= a_{11}x_1 + a_{12}x_2 + \dots + a_{1p}x_p \\y_2 &= a_{21}x_1 + a_{22}x_2 + \dots + a_{2p}x_p \\&\dots \\y_p &= a_{p1}x_1 + a_{p2}x_2 + \dots + a_{pp}x_p\end{aligned}$$

Переменные y_1, y_2, \dots, y_p называются главными компонентами или факторами. Таким образом, **фактор – это искусственный статистический показатель, возникающий в результате специальных преобразований корреляционной матрицы.**



7. ФАКТОРНЫЙ АНАЛИЗ

Процедура извлечения факторов называется факторизацией матрицы. В результате факторизации из корреляционной матрицы может быть извлечено разное количество факторов, вплоть до числа, равного количеству исходных переменных. Однако факторы, определяемые в результате факторизации, как правило, не равноценны по своему значению.



7. ФАКТОРНЫЙ АНАЛИЗ

Коэффициенты a_{ij} (факторная нагрузка), определяющие новую переменную, выбираются таким образом, чтобы новые переменные (главные компоненты, факторы) описывали максимальное количество вариативности данных (дисперсии) и не коррелировали между собой. Часто полезно представить коэффициенты a_{ij} таким образом, чтобы они представляли собой коэффициент корреляции между исходной переменной и новой переменной (фактором). Это достигается умножением a_{ij} на стандартное отклонение фактора. В большинстве статистических пакетов так и делается (в программе STATISTICA тоже).

7. ФАКТОРНЫЙ АНАЛИЗ

Коэффициенты называются факторными нагрузками. Обычно они представляются в виде таблицы, где факторы располагаются в виде столбцов, а переменные в виде строк. Такая таблица называется **таблицей факторных нагрузок** (матрицей) a_{ij} .

Переменные	Фактор 1	Фактор 2
органические удобрения	0,86	0,02
минеральные удобрения	0,75	0,01
известь	0,91	0,18
пестициды	0,34	0,35
содержание гумуса	0,13	0,85
кислотность почв	0,21	0,82

7. ФАКТОРНЫЙ АНАЛИЗ

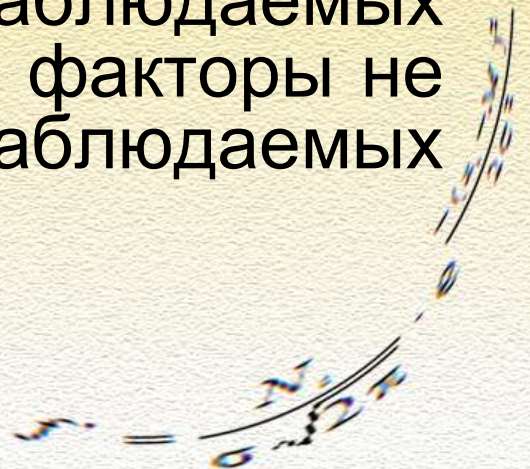
Числа, приведенные в ней, являются коэффициентами a_{ij} . Число 0,86 означает, что корреляция между первым фактором и органическими удобрениями равна 0,86. Чем выше факторная нагрузка по абсолютной величине, тем сильнее связь переменной с фактором.

Переменные	Фактор 1	Фактор 2
органические удобрения	0,86	0,02
минеральные удобрения	0,75	0,01
известь	0,91	0,18
пестициды	0,34	0,35
содержание гумуса	0,13	0,85
кислотность почв	0,21	0,82



7. ФАКТОРНЫЙ АНАЛИЗ

Факторный анализ отличается от метода главных компонент тем, что в его основе лежит предположение о некотором небольшом количестве фундаментальных переменных, которые не могут быть измерены прямо. Основное отличие между факторным анализом и методом главных компонент заключается в том, что главные компоненты являются линейными функциями от наблюдаемых переменных, в то время как общие факторы не выражаются через комбинацию наблюдаемых переменных.



7. ФАКТОРНЫЙ АНАЛИЗ

Модель факторного анализа предполагает, что корреляции между наблюдаемыми переменными x_1, x_2, \dots, x_p получаются благодаря их связи с некоторыми фундаментальными переменными, известными как общие факторы или латентные переменные f_1, f_2, \dots, f_k , где $k < p$.

В математической записи модель факторного анализа выглядит так:

$$\begin{aligned}x_1 &= \lambda_{11}f_1 + \lambda_{12}f_2 + \dots + \lambda_{1k}f_k + u_1 \\x_2 &= \lambda_{21}f_1 + \lambda_{22}f_2 + \dots + \lambda_{2k}f_k + u_2 \\&\dots \\x_p &= \lambda_{p1}f_1 + \lambda_{p2}f_2 + \dots + \lambda_{pk}f_k + u_p\end{aligned}$$

Случайная погрешность u_i называется характерностью и представляет собой часть наблюдаемой переменной, которая не объясняется действием факторов.

7. ФАКТОРНЫЙ АНАЛИЗ

Если латентные факторы не коррелируют, то коэффициенты λ_{ij} являются корреляциями между латентными переменными и явными переменными. Они также называются **факторными нагрузками** и представляются в виде такой же таблицы, как и факторные нагрузки в методе главных компонент.


Соответствие факторной модели полученным данным проверяется путем сравнения исходной корреляционной матрицы с матрицей корреляций, полученной в результате применения модели. Такая оценка соответствия может быть проведена различными методами, которые носят название ***principal factor analysis*** (анализ главных факторов).



7. ФАКТОРНЫЙ АНАЛИЗ

Сколько факторов следует выделять?

Поскольку факторный анализ является методом сокращения числа переменных. Отсюда в каждом исследовании вопрос: сколько факторов следует выделять? В процессе последовательного выделения факторов они включают в себя все меньше и меньше изменчивости (то есть объясняют все меньше и меньше дисперсии). Решение о том, когда следует остановить процедуру выделения факторов, главным образом зависит от точки зрения на то, что считать малой «случайной» изменчивостью. Это решение достаточно произвольно, однако имеются некоторые рекомендации, позволяющие рационально выбрать число факторов.



7. ФАКТОРНЫЙ АНАЛИЗ

Для применения процедуры выбора следует посчитать некоторую статистику – собственные значения корреляционной матрицы и процент объясненной дисперсии для каждого фактора. Собственное значение – это характеристика матрицы корреляций, которая используется для декомпозиции матрицы и одновременно как критерий определения числа выделенных факторов и как мера дисперсии, соответствующей данному фактору. Если разделить собственное значение на число переменных p , то получится доля дисперсии, соответствующая данному фактору. Посчитаем эти показатели для примера с агроландшафтом.

7. ФАКТОРНЫЙ АНАЛИЗ

В данной таблице шесть факторов – столько, сколько и было переменных в исследовании. Больше взять нельзя. Так сколько же факторов нужно отбирать?

Фактор	Собственные значения			
	Собственные значения	% общей дисперсии	Кумулят. соб. знач.	Кумулят. % общей дисперсии
1	3,811753	63,52922	3,811753	63,5292
2	1,096612	18,27686	4,908365	81,8061
3	,416857	6,94762	5,325222	88,7537
4	,295090	4,91817	5,620312	93,6719
5	,197790	3,29651	5,818103	96,9684
6	,181897	3,03162	6,000000	100,0000

Value	Eigenvalues (factor) Extraction: Principal components			
	Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative %
1	2,893561	36,16951	2,893561	36,16951
2	1,708925	21,36156	4,602486	57,53108

7. ФАКТОРНЫЙ АНАЛИЗ

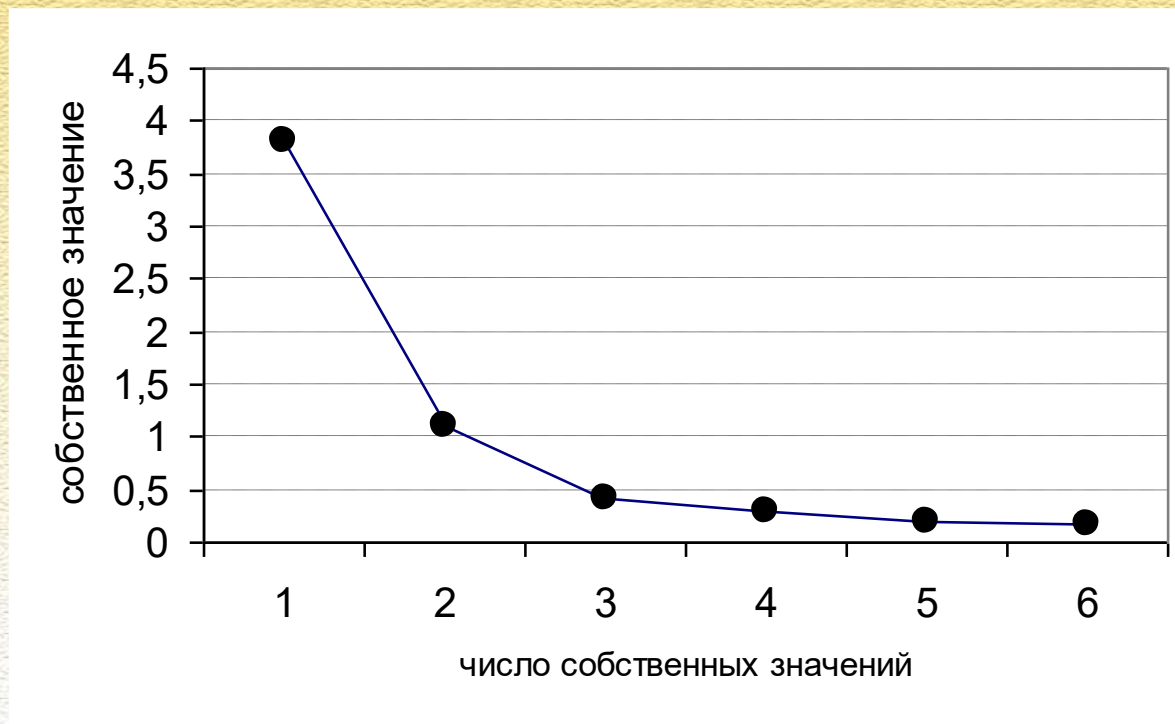
Это можно сделать на основании следующих критериев:

1. Процент объясненной дисперсии. Если кумулятивный (накопленный) процент общей дисперсии достигает 60% или больше, то можно остановиться на данном количестве факторов. Из приведенной ранее таблицы видно, что достаточно взять даже один фактор.

2. Критерий Кайзера (H. Keiser). Отбираются только факторы с собственными значениями, большими 1. По существу, это означает, что если фактор не выделяет дисперсию, эквивалентную, по крайней мере, дисперсии одной переменной, то он опускается. Этот критерий предложен Кайзером, и является, вероятно, наиболее широко используемым. В приведенном выше примере на основе этого критерия вам следует сохранить только 2 фактора (две главные компоненты).

7. ФАКТОРНЫЙ АНАЛИЗ

3. Критерий каменной осыпи – графический метод. Изобразим собственные значения, представленные в таблице ранее, в виде простого графика. Следует найти такое место на графике, где убывание собственных значений слева направо максимально замедляется. Предполагается, что справа от этой точки находится только «факториальная осыпь». В соответствии с этим критерием можно оставить в этом примере 2 или 3 фактора.



7. ФАКТОРНЫЙ АНАЛИЗ

Критерий Кайзера иногда сохраняет слишком много факторов, в то время как критерий каменистой осыпи иногда сохраняет слишком мало факторов; однако оба критерия вполне хороши при нормальных условиях, когда имеется относительно небольшое число факторов и много переменных.

В ряде исследований распределение переменных по факторам не всегда бывает ясным и простым. Поэтому обычно исследуется несколько решений с бóльшим или меньшим числом факторов, и затем выбирается одно наиболее «осмысленное».

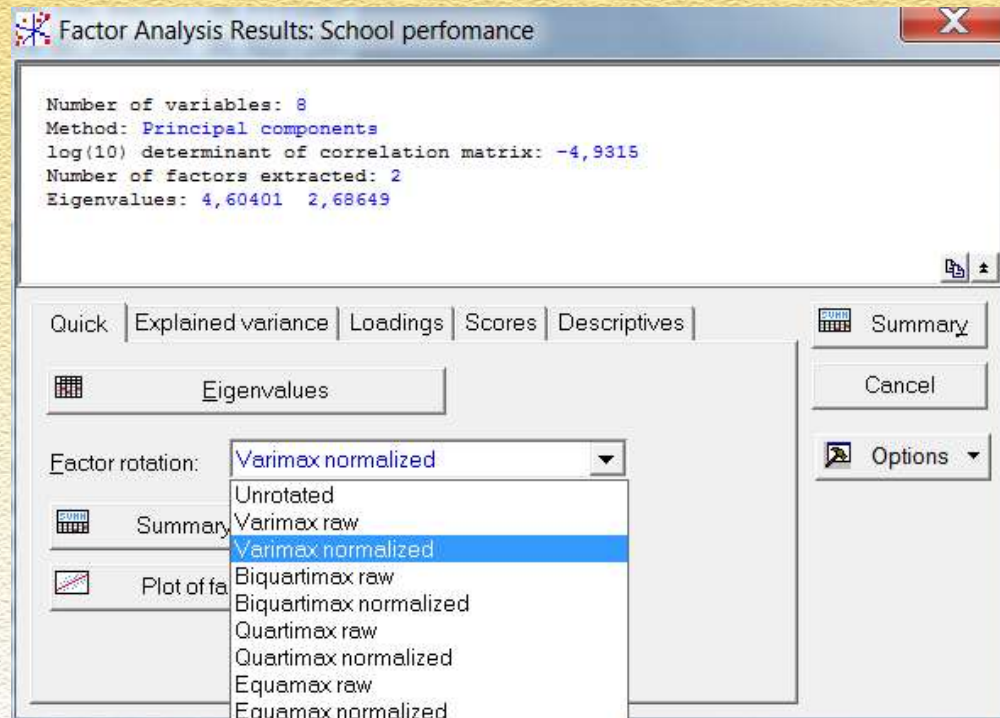
7. ФАКТОРНЫЙ АНАЛИЗ

В нашем примере с агроландшафтом все довольно просто. Факторная матрица показывает, какие переменные образуют каждый фактор. Это связано, прежде всего, с абсолютным значением факторной нагрузки. Программа STATISTICA по умолчанию выбирает минимальное значение факторной нагрузки равное 0,7 (допускается только сильная связь между переменной и фактором).

Переменные	Фактор 1	Фактор 2
органические удобрения	0,86	0,02
минеральные удобрения	0,75	0,01
известь	0,91	0,18
пестициды	0,34	0,35
содержание гумуса	0,13	0,85
КИСЛОТНОСТЬ ПОЧВ	0,21	0,82

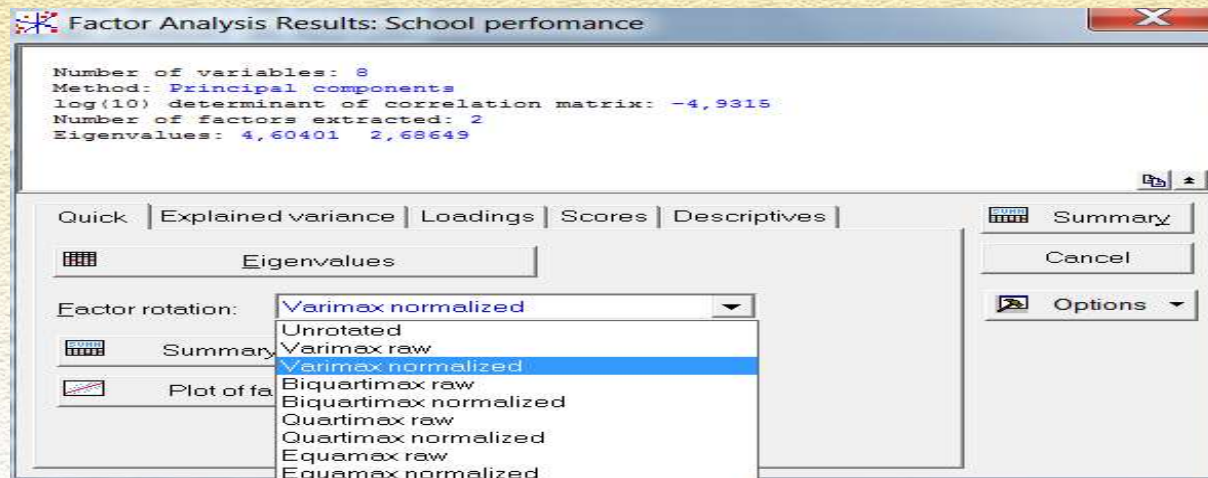
7. ФАКТОРНЫЙ АНАЛИЗ

Иногда оказывается, что описанные выше шаги не дают однозначного решения задачи определения факторов. Основываясь на геометрическом представлении рассматриваемой задачи, поиск однозначного решения называют задачей вращения факторов. Необходимость вращения факторов возникает чаще всего, когда выявленным факторам не удастся дать достаточно четкую содержательную интерпретацию.



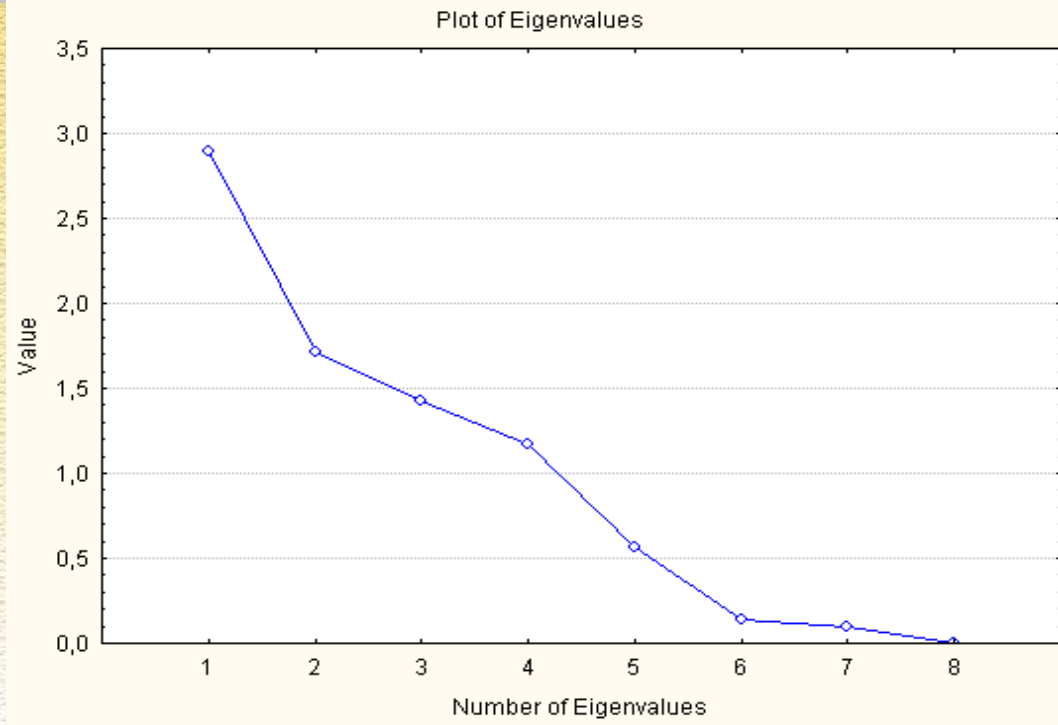
7. ФАКТОРНЫЙ АНАЛИЗ

Например, факторные нагрузки для рассматриваемого фактора могут быть близкими по величине и одинаковыми по знаку у многих признаков, так что трудно однозначно определить, какой фактор «стоит» за выделенной комбинацией признаков. Вращение позволяет сделать матрицу факторных нагрузок более «контрастной» за счет увеличения нагрузок по одним признакам и уменьшения по другим, что способствует более отчетливому выявлению групп признаков, определяющих тот или иной фактор. Факторные нагрузки повернутой матрицы можно рассматривать как результаты выполнения процедуры факторного анализа. Кроме того, на основании значений этих нагрузок необходимо попытаться дать толкование отдельным факторам.



7. ФАКТОРНЫЙ АНАЛИЗ

Eigenvalues (factor)				
Extraction: Principal components				
Value	Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative %
1	2,893561	36,16951	2,893561	36,16951
2	1,708925	21,36156	4,602486	57,53108

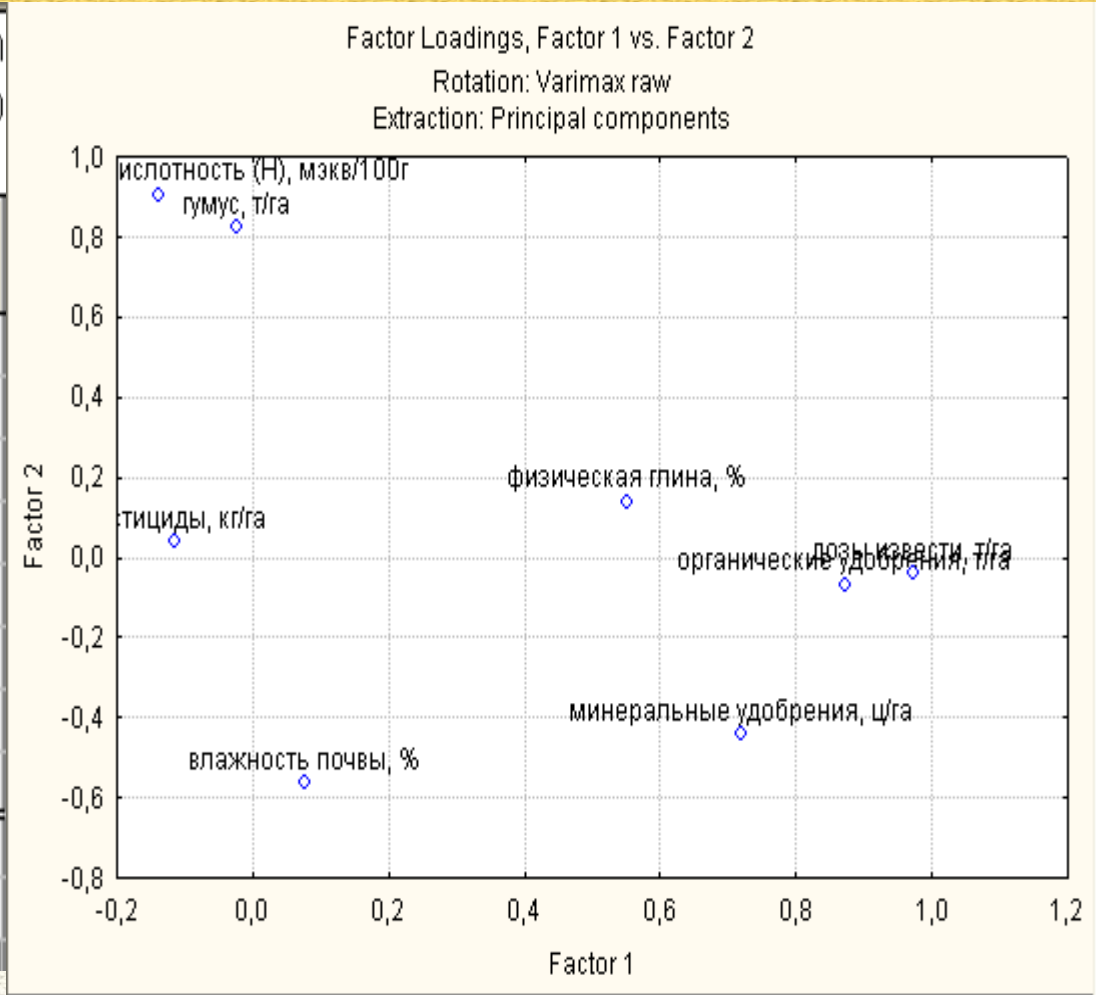


Результаты факторного анализа показывают, что необходимо выделять как минимум 3 фактора



7. ФАКТОРНЫЙ АНАЛИЗ

Variable	Factor Loadings (Varimax)	
	Factor 1	Factor 2
органические удобрения, т/га	0,872127	-0,067592
минеральные удобрения, ц/га	0,718471	-0,441738
дозы извести, т/га	0,972194	-0,040124
пестициды, кг/га	-0,115877	0,040150
гумус, т/га	-0,022341	0,820951
гидролитическая кислотность (Н), мэкв/100г	-0,138428	0,906199
влажность почвы, %	0,075795	-0,565191
физическая глина, %	0,552192	0,138743
Expl. Var	2,566717	2,036770
Prp. Totl	0,320715	0,254596





Контрольные вопросы по лекции

Контрольные вопросы:

- Какие условия применения факторного анализа?
 - С помощью каких критериев определяют количество выделяемых факторов?
 - Что показывает факторная нагрузка?
 - Зачем применяют «вращение» факторной матрицы?
 - Что показывает процент объясненной дисперсии?
- 