

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ
БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ»

Ю.М. Фетисов

**МЕТОДЫ ФАКТОРНОГО АНАЛИЗА
И ШКАЛИРОВАНИЯ
В ГЕОГРАФИИ И ГЕОЭКОЛОГИИ**

Учебно-методическое пособие для вузов

Издательско-полиграфический центр
Воронежского государственного университета
2014

ПРЕДИСЛОВИЕ

Наиболее содержательные разделы статистического исследования связаны с многомерным анализом данных. Прикладные методы многомерной статистики широко используются в современных эколого-географических исследованиях. В отличие от одномерной выборки в многомерном анализе элемент системы описывается множеством переменных. Измерение сразу нескольких признаков (свойств объекта) в одном эксперименте более естественно, чем измерение лишь какого-то одного. Та часть математической статистики, которая исследует эксперименты с такими многомерными наблюдениями, называется многомерным статистическим анализом.

В STADIA содержится большой набор самых распространенных процедур многомерного анализа. Их грамотное использование требует от географа и геоэколога как знания по теории многомерного статистического анализа, так и владения навыками практической работы в среде этого пакета.

Пособие содержит начальные сведения о пакете STADIA, позволяющие освоить основные приемы работы с массивами данных, а так же научиться практическому использованию двух модулей пакета — факторного анализа и шкалирования.

Для успешного освоения каждого из этих модулей вначале даётся краткий теоретический материал, а затем рассматривается типовая задача с указанием всех пунктов меню, приводящих к ее решению. С целью проверки знаний и закрепления навыков в пособии приведены задания для самостоятельной работы. Такая структура оказывается очень удобной при приеме зачетов по каждой теме, а так же при использовании пособия для самостоятельного изучения.

4. *Диалог.* Далее протекает диалог, характерный для выполняемого метода с выдачей числовых результатов анализа и их интерпретации в экранную страницу [Rez] текстового редактора, а графиков результатов – в графические страницы [Gr_i], i=1,...15.

Текстовый редактор становится доступным при активизации страницы результатов анализа [Rez]. Он поддерживает большинство типичных для всех подобных редакторов операций: ввод текста с клавиатуры, выделение фрагментов текста, удаление символов (только клавишей Back Space) и фрагментов, забор фрагментов в буфер обмена и вставление фрагментов из буфера. Редактор поддерживает также общие операции: чтение и запись в отношении текстовых файлов, изменение шрифта и выдачи результатов на печать. Операция *чтения* полезна для компоновки общего отчета о полученных в разное время результатах. По такой операции содержимое страницы результатов полностью замещается. Поэтому перед выполнением чтения полезно забрать в буфер имеющуюся выдачу.

Для переноса числовых результатов анализа в электронную таблицу необходимо забирать их в *буфер обмена* из страницы результатов, перейти в электронную таблицу и в нужном месте вставить содержимое буфера обмена.

Полученные в графической форме результаты могут быть перенесены в электронную таблицу посредством специальной инструментальной клавиши «*СохрГраф.*». Координаты точек каждого графика переносятся в первые свободные столбцы электронной таблицы.

Система STADIA выпускается в четырех модификациях (учебная, студенческая, базовая и профессиональная), отличающихся только объемом обрабатываемых данных (соответственно 400, 4000, 20000 и 32000 чисел совокупно в матрице данных). Учебная STADIA с файлами примеров свободно доступна по адресу: <http://statsoft.msu.ru/Stadia.zip>. Это дает возможность студентам использовать данный пакет вне учебных компьютерных классов.

STADIA постоянно развивается, появляются более совершенные версии этой системы, открывая новые возможности для пользователей.

2. ФАКТОРНЫЙ АНАЛИЗ

2.1. Теоретические основы

Факторный анализ — основной метод изучения структуры взаимосвязей между переменными (классификация переменных) и проявления внутренних закономерностей путем «сжатия» статистической информации. Поэтому факторный анализ используется или как метод классификации, или как метод сокращения данных.

Основной задачей факторного анализа является переход от первоначальной системы большого числа взаимосвязанных показателей x_1, x_2, \dots, x_m к относительно малому числу скрытых (латентных) факторов $f_1, f_2, \dots, f_k, k < m$. Например, если почвы ряда районов охарактеризованы комплексом признаков по влажности почвы, содержание гумуса в пахотном слое, реакции среды, объемом вносимых удобрений и ядохимикатов и т.д., то результатом факторного анализа могут быть «фактор естественного плодородия почвы», «фактор сельскохозяйственной химизации» и т.д.

Пусть имеется n объектов, каждый из которых описывается m признаками. Например, объекты — это районы Воронежской области, а признаки — это факторы, определяющие развитие негативных природных процессов в этих районах. Представим исходную информацию в виде матрицы $X = \{x_{ij}\}$ размерности $n \times m$. Так как массив данных $X = \{x_{ij}\}$ представляет величины различной размерности, то для того чтобы перейти к безразмерным величинам, пронормируем элементы массива:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j},$$

где x_{ij} — значение j -го признака i -го объекта; \bar{x}_j — выборочное среднее j -й переменной (признака); s_j — среднее квадратическое отклонение j -го переменной. Получим матрицу нормированных значений $Z=\{z_{ij}\}$. Предполагается, что каждый элемент этой матрицы z_{ij} является результатом воздействия некоторого числа k ($k \ll m$) гипотетических общих факторов и одного характерного фактора:

$$z_{ij} = \sum_{v=1}^k a_{iv} f_{iv} + d_j \varepsilon_{ij}, \quad (2.1)$$

где a_{iv} — неизвестные коэффициенты, называемые факторными нагрузками;

f_{iv} — значение i -го общего фактора на v -м объекте исследования;

d_j — нагрузки первоначальных признаков на характерные факторы;

ε_{ij} — характерные факторы.

Первый член правой части равенства (2.1) показывает долю первого фактора в исследуемых явлениях, второй — долю второго фактора и т.д., последний — долю независимого фактора (остаток).

Особенностью факторного анализа является неоднозначность определения общих факторов. Существует несколько вариантов факторного анализа, причем наиболее часто применяемым вариантом служит компонентный анализ или метод главных компонент. Компонентный анализ можно рассматривать как частный случай метода факторного анализа, когда все специальные факторы приняты равными нулю, а общие факторы ортогональны. Анализ главных компонент более предпочтителен как метод сокращения данных, а анализ факторов целесообразнее применять для определения структуры данных.

В отличие от рассматриваемых в факторном анализе общих факторов, которые определяют большую (но не всю) часть вариации первоначальных признаков, главные компоненты объясняют всю вариацию и определяются однозначно. Модель главных компонент имеет вид: