

АЛГОРИТМ МИНИМИЗАЦИИ КОЛИЧЕСТВА ПАРАМЕТРОВ ИССЛЕДОВАНИЯ В ГЕОЭКОЛОГИИ

На базе факторного и регрессионного анализов предлагается минимизация числа параметров исследования для многопараметрического объекта: окружающая среда и новообразования. За счет выбора базовых параметров значительно сужается пространство параметров без ухудшения точностных характеристик описания многопараметрического объекта.

Предлагается следующий алгоритм минимизации числа параметров исследования для многопараметрических объектов:

1. Строим матрицу исследования (строчки – наблюдения, столбцы – параметры исследования).

2. Методом главных компонент находим матрицу факторных нагрузок. Осуществляем вариационное вращение в пространстве факторов (строчки в матрице факторных нагрузок – параметры исследования, столбцы – гипотетические переменные, факторы) [3, 4].

3. В каждой строчке матрицы факторных нагрузок, то есть для каждого параметра исследования, находим максимальную по модулю факторную нагрузку.

4. Определяем по каждому фактору попадание в этот фактор параметров с максимальной по модулю факторной нагрузкой (пункт 3). То есть тем самым определяем объединение параметров по факторам.

5. В объединившихся в каждом факторе параметрах выбираем один параметр с максимальной по модулю факторной нагрузкой. Число таких выбранных параметров будет равно, очевидно, числу факторов.

6. Строим для всех параметров исследования полиномиальные модели, аргументами в которых будут выбранные в пункте 5 параметры [1, 2, 5].

7. По построенным моделям для каждого параметра осуществляем определение вкладов параметров-аргументов (оценку количественной обусловленности параметров выбранными параметрами).

8. Сравниваем качественные групповые обусловленности, объединения параметров по факторам, с количественными обусловленностями параметров, полученными в пункте 7.

Если групповые и количественные обусловленности для всех параметров исследования не будут сильно отличаться по числу несовпадений, то выбранные в пункте 5 параметры могут быть приняты за базисные при описании данного многопараметрического объекта, матрица исследования которого была взята за основу в данном алгоритме. То есть тем самым осуществляем минимизацию количества параметров исследования, потому что число факторов меньше числа параметров.

Рассмотрим использование разработанного алгоритма для матрицы исследования с параметрами-столбиками: смертность от новообразований и окружающая среда по Оренбургской области.

Согласно алгоритму осуществляем выбор базисных параметров:

в факторе-2 базовый параметр – 15 (сажа при сгорании бензина в % по объему);

в факторе-4 базовый параметр – 51 (анилин в воздухе, мг/м³);

в факторе-6 базовый параметр – 59 (кобальт металлический в воздухе мг/м³);

в факторе-3 базовый параметр – 25 (акриловая кислота в воде, мг/л);

в факторе-7 базовый параметр – 17 (хлорамп (арборицид), мг/кг почвы);

в факторе-5 базовый параметр – 34 (кобальт в воде, мг/л).

Согласно факторному анализу (таблица 1) параметры 4, 6 имеет групповую обусловленность с параметром 15.

Согласно вкладам параметры 4, 6 больше всего обусловлены параметром 15 (таблицы 2, 4).

Результаты факторного анализа (фрагмент)

Таблица 1. Объединение по фактору 2

НОМЕР	НАЗВАНИЕ ПАРАМЕТРА	НАГРУЗКА
4	(рак матки в чел. умерших)	.7535
6	(рак пищевода в чел. умерших)	.7266
12	(оксиды азота при сгорании бензина в % по объему)	-.8558
14	(альдегиды при сгорании бензина в % по объему)	-.6500
15	(сажа при сгорании бензина в % по объему)	-.9642
21	(гамма-изомер гексахлорана (инсектицид) мг/кг почвы)	.9173
46	(цинк в воде мг/л)	-.7655
49	(аммиак в воздухе мг/м ³)	-.6437
57	(кислота серная в воздухе мг/м ³)	.6992
63	(хлор в воздухе мг/м ³)	.5681

Промечание: в факторе 2, базовый параметр – 15 (сажа при сгорании бензина в % по объему).

Результаты регрессионного анализа (фрагмент)

Зависимый параметр – 4
(рак матки в чел. умерших)

Таблица 2. Вклады параметров-аргументов в модели

НОМЕР	НАЗВАНИЕ ПАРАМЕТРА	ВКЛАД В МОДЕЛЬ
15	(сажа при сгорании бензина в % по объему)	1.0224
51	(анилин в воздухе мг/м3)	-0.0852
59	(кобальт металлический в воздухе м/г/м3)	0.0265
25	(акриловая кислота в воде мг/л)	0.0000
17	(хлорамп (арборицид) мг/кг почвы)	0.0263
34	(кобальт в воде мг/л)	0.0100

Модель (рак матки в чел. умерших) значимый по вкладу параметр – 15 (сажа при сгорании бензина в % по объему) вклад = 1.0223810

$$y = + (234.3866) * (x 15)^{**0} + (-1300.7100) * (x 15)^{**1} + (-17662.4300) * (x 15)^{**2} + (-634.6078) * (x 51)^{**0} + (45451.6500) * (x 51)^{**1} + (-960005.3000) * (x 51)^{**2} + (4835900.0000) * (x 51)^{**3} + (41.3868) * (x 59)^{**0} + (-36600.7700) * (x 59)^{**1} + (-28.7673) * (x 17)^{**0} + (651.7759) * (x 17)^{**1} + (-18.7846) * (x 34)^{**0} + (18.6961) * (x 34)^{**1}$$

* - умножение, ** - возведение в степень

Таблица 3. Характеристики модели

ХАРАКТЕРИСТИКИ МОДЕЛИ	ЗНАЧЕНИЯ
Коэффициент детерминации	.95
Средняя абсолютная ошибка	4.37
Средняя ошибка в процентах	3.03

То есть здесь мы имеем совпадение групповых и количественных обусловленностей для зависимых параметров 4, 6 и параметра-аргумента 15.

Аналогичный анализ групповых и количественных обусловленностей был проведен для остальных параметров исследования.

Количественная обусловленность параметров совпала с групповой обусловленностью для всех параметров.

Зависимый параметр – 6
(рак пищевода в чел. умерших)

Таблица 4. Вклады параметров-аргументов в модели

НОМЕР	НАЗВАНИЕ ПАРАМЕТРА	ВКЛАД В МОДЕЛЬ
15	(сажа при сгорании бензина в % по объему)	.9755
51	(анилин в воздухе мг/м3)	0.0212
59	(кобальт металлический в воздухе м/г/м3)	0.0070
25	(акриловая кислота в воде мг/л)	0.0165
17	(хлорамп (арборицид) мг/кг почвы)	-0.0091
34	(кобальт в воде мг/л)	-0.0112

Модель (рак пищевода в чел. умерших) значимый по вкладу параметр – 15 (сажа при сгорании бензина в % по объему) вклад = .9755280

$$y = + (350.1800) * (x 15)^{**0} + (-5068.2570) * (x 15)^{**1} + (2580.2430) * (x 15)^{**2} + (527.0512) * (x 51)^{**0} + (-42413.2400) * (x 51)^{**1} + (1107744.0000) * (x 51)^{**2} + (-9386275.0000) * (x 51)^{**3} + (-72.9595) * (x 59)^{**0} + (343249.8000) * (x 59)^{**1} + (-453330100.0000) * (x 59)^{**2} + (183648600000.0000) * (x 59)^{**3} + (-10.9827) * (x 25)^{**0} + (-72.8374) * (x 25)^{**1} + (189.7899) * (x 25)^{**2} + (9.4853) * (x 17)^{**0} + (-216.0154) * (x 17)^{**1} + (-71.4056) * (x 34)^{**0} + (64.5448) * (x 34)^{**1}$$

* - умножение, ** - возведение в степень

Таблица 5. Характеристики модели

ХАРАКТЕРИСТИКИ МОДЕЛИ	ЗНАЧЕНИЯ
Коэффициент детерминации	.95
Средняя абсолютная ошибка	4.75
Средняя ошибка в процентах	3.36

Таким образом, данный многопараметрический объект можно описать шестью базисными параметрами вместо исходных 66. Остальные параметры можно определять по регрессионным моделям на шести базисных параметрах-аргументах. Все эти модели имеют хорошие характеристики (таблицы 3, 5).

Разработанный метод может с успехом использоваться во многих областях исследований, связанных с многопараметрическими объектами.

Список использованной литературы:

1. Бендат Д.Ж., Пирсол А. Измерение и анализ случайных процессов. – М.: «Мир», 1974.
2. Драйпер Н., Смит Г. Прикладной регрессионный анализ. – М.: Статистика, 1993.
3. Иберла К. Факторный анализ. – М.: Статистика, 1980.
4. Харман Г. Современный факторный анализ. – М.: Статистика, 1972.
5. Brandon D.B. Developing Mathematical Models for Computer.