УДК 519.24

Л.Г. Раскин, О.В. Серая

Национальный технический университет «ХПИ», Харьков

ИНФОРМАЦИОННЫЕ ПРОБЛЕМЫ КАНОНИЧЕСКОГО РЕГРЕССИОННОГО АНАЛИЗА

Рассмотрены технологии решения задач регрессионного анализа для случаев, когда нарушаются традиционные исходные предпосылки. При этом разработаны методы оценивания параметров уравнения регрессии, если независимые переменные – случайные величины или заданы нечетко. Предложено решение задачи компараторной идентификации для нечетких исходных данных.

***Ключевые слова**: регрессионный анализ, нарушение предпосылок, нечеткие исходные данные, компараторная идентификация.*

Введение

Регрессионный анализ – мощный и эффективный статистический метод построения математических моделей, описывающих зависимость между показателем функционирования анализируемой системы *y* и обусловливающими, объясняющими независимыми переменными (факторами) $F_1, F_2, ..., F_m$. С целью выявления этой связи проводится серия экспериментов, в которой каждому опыту $\left(F_{j1}, F_{j2}, ..., F_{jm}\right)$ ставится в соответствие его результат – значение зависимой переменной y_j, $j = 1, 2, ..., n$. Искомая связь обычно описывается полиномом Колмогорова-Габора, который в простейшем случае имеет вид:

$$y_j = x_0 + F_{j1}x_1 + F_{j2}x_2 + ... + F_{jm}x_m + \varepsilon_j. \quad (1)$$

Здесь F_{ji} – значение i-й независимой переменной в j – м опыте, $i = 0, 1, 2, ..., m$, $j = 1, 2, ..., n$.

В матричной форме соотношение (1) имеет вид $FX = Y$,где

$$F = \begin{pmatrix} 1 & F_{11} & F_{12} & ... & F_{1m} \\ 1 & F_{21} & F_{22} & ... & F_{2m} \\ ... & ... & ... & ... & ... \\ 1 & F_{n1} & F_{n2} & ... & F_{nm} \end{pmatrix}, \; X = \begin{pmatrix} x_0 \\ x_1 \\ ... \\ x_m \end{pmatrix}, \; Y = \begin{pmatrix} y_1 \\ y_2 \\ ... \\ y_n \end{pmatrix}.$$

В каноническом регрессионном анализе делаются следующие основные предположения.

1. Значения независимых переменных F_i измеряются без ошибок, $i = 1, 2, ..., m$.

2. Зависимая переменная y в каждом опыте оценивается со случайной ошибкой ε_j, которая нормально распределена с нулевым математическим ожиданием и известной дисперсией σ^2.

3. Случайные ошибки ε_j в разных опытах не коррелированы.

В этих предложениях оценки неизвестных коэффициентов $x_0, x_1, ..., x_m$ регрессионного полинома (1) получают методом наименьших квадратов (МНК), минимизируя сумму квадратов отклонений значений результирующей переменной y_j от соответствующих значений $\sum_{i=0}^{m} x_i F_{ji}$, предсказываемых моделью (1).

Постановка задачи. При решении многих практических задач возникают ситуации, когда исходные предпосылки классического регрессионного анализа не верны. При этом, естественно, шаблонное применения МНК может привести к грубым ошибкам. Поставим задачу разработки методик оценивания коэффициентов уравнения регрессии для разных вариантов нарушений канонических предпосылок.

Основные результаты

А. Оценивание параметров регрессионной модели по данным экспериментов со случайными ошибками в значениях независимых переменных (нарушена предпосылка 1).

Будем считать, что наблюдаемое значение F_{ji} есть нормально распределенная случайная величина с математическим ожиданием, равным истинному значению w_{ji}, и дисперсией σ_{ji}^2, а наблюдаемое значение Y_j есть нормально распределенная случайная величина с математическим ожиданием, равным истинному значению ν_j, и дисперсией σ_j^2.

При этом, как легко показать [1], оценки регрессионных коэффициентов, получаемых с использованием МНК, оказываются смещенными. В [2] доказано, что величина смещения зависит от исходных данных и может быть очень большой. Известные методы снижения уровня смещения [3,4], сопряжены с необходимостью проведения сложных вычислений и приводят к приемлемым результатам, если ошибки оценивания независимых переменных малы.

В связи с этим рассмотрим другой подход к решению задачи оценивания регрессионных коэф-

фициентов в условиях, когда значения независимых переменных оцениваются с ошибкой.

Принятые предположения с учетом принципа инверсии позволяют записать законы распределения неизвестных истинных значений независимых и зависимой переменных в виде

$$\mu\left(w_{ji}\right) = \frac{1}{\sqrt{2\pi}\sigma_{ji}} \exp\left\{-\left(w_{ji} - F_{ji}\right)^2 \Big/ \left(2\sigma_{ji}{}^2\right)\right\}, \quad (2)$$

$$\mu\left(v_j\right) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left\{-\left(v_j - y_j\right)^2 \Big/ \left(2\sigma_j^2\right)\right\},$$

$$i = 1, 2, ..., m, \;\; j = 1, 2, ..., n. \quad (3)$$

Если модель (1) адекватна, то для истинных значений наблюдаемых величин должны выполняться равенства:

$$\begin{aligned} &x_0 + w_{11}x_1 + w_{12}x_2 + ... + w_{1m}x_m = v_1, \\ &..... \\ &x_0 + w_{n1}x_1 + w_{n2}x_2 + ... + w_{nm}x_m = v_n. \end{aligned} \quad (4)$$

Поскольку параметры системы (4) – случайные величины, то соотношения (2) - (4) задают стохастическую систему линейных алгебраических уравнений. Теперь поступаем следующим образом. Решим порождаемую (4) детерминированную систему линейных алгебраических уравнений, заменив случайные параметры $\left(w_{ji}\right)$ и $\left(v_i\right)$ их математическими ожиданиями. При этом получим

$$\sum_{i=0}^{m} F_{ji}x_i = y_j, \;\; j = 1, 2, ..., n. \quad (5)$$

Эта переопределенная система в матричной форме имеет вид

$$FX = Y, \;\; F = \left(F_{ji}\right),$$

$$F_{jo} = 1, i = 0, 1, 2, ..., m, \; j = 1, 2, ..., n,$$

решается путем минимизации функционала

$$J = \left(FX - Y\right)^T \left(FX - Y\right),$$

а искомый вектор – решение имеет вид

$$X^{(0)} = \left(F^T F\right)^{-1} F^T Y. \quad (6)$$

Назовем это решение модальным.

Введем далее набор переменных $Z_j(X)$:

$$\begin{aligned} &Z_1(X) = w_{11}x_1 + w_{12}x_2 + ... + w_{1m}x_m, \\ &..... \\ &Z_n(X) = w_{n1}x_1 + w_{n2}x_2 + ... + w_{nm}x_m, \end{aligned} \quad (7)$$

задающих предсказываемые моделью значения зависимой переменной в каждом опыте. Введенные переменные – это случайные величины, плотности распределения которых совместно с $\left(\mu(w_{ij})\right)$ определяются набором X:

$$\phi(Z_j) = \frac{1}{\sqrt{2\pi D_j}} \exp\left\{-\frac{(Z_j - m_j)^2}{2D_j}\right\},$$

$$m_j = \sum_{i=0}^{m} F_{ji}x_i, \;\; D_j = \sum_{i=0}^{m} \sigma_{ji}^2 x_i^2, . \quad (8)$$

$$\sigma_{j0}^2 = 0, \; j = 1, 2, ..., n.$$

Теперь детерминированным решением стохастической задачи (4) будем называть набор $X = (x_i), i = 1, 2, ..., m,$ минимизирующий сумму дисперсий случайных величин $Z_j(x)$ и наименее уклоняющийся от $X^{(0)}$. Смысл этого критерия понятен. Его использование обеспечивает получение набора детерминированных оценок регрессионных коэффициентов $x_1, x_2, ..., x_m,$ для которых плотности распределения случайных величин $Z_1, Z_2, ..., Z_n$, определяющих предсказываемые моделью значения зависимой переменной в каждом опыте, наименее размыты, и имеющих математические ожидания, максимально близкие к нулю. Формальное описание критерия имеет вид:

$$\begin{aligned} L(x) &= \sum_{j=1}^{n} D_j(x) + \sum_{i=0}^{m} (x_i - x_i^{(0)})^2 = \\ &= \sum_{i=0}^{m} \left[x_i^2 \left(\sum_{j=1}^{n} \sigma_{ji}^2 \right) + \left(x_i - x_i^{(0)} \right)^2 \right] = \\ &= \sum_{i=0}^{m} \left[x_i^2 D_{\Sigma i} + (x_i - x_i^{(0)})^2 \right]. \end{aligned} \quad (9)$$

Найдем набор X, минимизирующий (9), методом множителей Лагранжа, дополнив модель задачи условием нормировки

$$\sum_{i=0}^{m} x_i = 1.$$

Запишем функцию Лагранжа

$$\Phi(x) = L(x) - \lambda\left(\sum_{i=0}^{m} x_i - 1\right).$$

Далее имеем

$$\frac{d\Phi(x)}{dx_i} = 2x_i(D_{\Sigma i} + 1) - 2x_i^{(0)} - \lambda = 0,$$

откуда

$$x_i = \frac{2x_i^{(0)} + \lambda}{2(D_{\Sigma_i} + 1)} = \frac{\lambda}{2} \cdot \frac{1}{D_{\Sigma_i} + 1} + \frac{x_i^{(0)}}{D_{\Sigma_i} + 1}.$$

Значение $\lambda/2$ найдем из условия нормировки

$$\sum_{i=0}^{m} x_i = \frac{\lambda}{2} \sum_{i=0}^{m} \frac{1}{D_{\Sigma i} + 1} + \sum_{i=0}^{m} \frac{x_i^{(0)}}{D_{\Sigma i} + 1} = 1,$$

$$\frac{\lambda}{2} = \left(1 - \sum_{i=0}^{m} \frac{x_i^{(0)}}{D_{\Sigma i} + 1}\right) \Big/ \sum_{i=0}^{m} \frac{1}{D_{\Sigma i} + 1}.$$

Тогда

$$x_i = \frac{1 - \sum_{i=0}^{m} \frac{x_i^{(0)}}{D_{\Sigma i} + 1}}{\sum_{i=0}^{m} \frac{1}{D_{\Sigma i} + 1}} \cdot \frac{1}{D_{\Sigma i} + 1} + \frac{x_i^{(0)}}{D_{\Sigma i} + 1}, \qquad (10)$$

$$i = 0, 1, ..., m.$$

Более корректный результат будет получен, если в соотношении (9) нормировать каждое слагаемое его максимальным значением.

Понятно, что если измерения проводятся без ошибок, то из (10), следует

$$x_i = x_i^{(0)}, i = 0, 1, ..., m.$$

Б. Оценивание параметров регрессионной модели по нечетким исходным данным.

Общая схема решения задачи в этом случае повторяет предыдущую. Так же, как и ранее, на первом шаге составляется и решается система уравнений (5), параметрами которой являются модальные значения нечетких чисел, описывающих исходные данные задачи. Соответствующее модальное решение системы получаем по формуле (6). Затем с использованием (7) задаются предсказываемые этой моделью значения зависимой переменной в каждом опыте. Теперь по правилам выполнения операций над нечеткими числами [5]определяются функции принадлежности получаемых при этом нечетких чисел, по которым рассчитываются уровни их размытости, например, по длине интервалов соответствующих носителей. Далее, по аналогии с (9) формируется составной критерий качества искомого решения.

В. Оценивание параметров регрессионной модели при отсутствии данных о значениях зависимой переменной. Компараторная идентификация.

В задачах отыскания связи между характеристиками объекта и его качеством(полезностью, эффективностью) непосредственное использование технологии регрессионного анализа затруднено, если при проведении опытов вся доступная информация ограничена только данными о значениях независимых переменных (F_{ij}).

Аналогичная ситуация возникает при сравнении совокупности однотипных объектов по данным о наборе их характеристик. Разумеется, этой информации недостаточно для построения регрессионной модели

$$R(x) = x_1 F_1 + x_2 F_2 + ... + x_m F_m, \qquad (11)$$

где $X = (F_1, F_2, ..., F_m)$ – набор факторов - характеристик объекта,

$X = (x_1, x_2, ..., x_m)$ – набор весовых коэффициентов.

Вместе с тем использование данных о характеристиках объектов позволяет, например, экспертным путем, ранжировать сравниваемые объекты по убыванию значения их результирующей характеристики. При этом получаем

$$R_1(X) > R_2(X) > ... > R_{n-1}(X) > R_n(X). \qquad (12)$$

Это соотношение используется в стандартной задаче компараторной идентификации [6] для отыскания коэффициентов уравнения (11). Неравенства (12) определяют следующую систему строгих неравенств:

$$\zeta_1(X) = R_2(X) - R_1(X) =$$

$$= x_1(F_{21} - F_{11}) + x_2(F_{22} - F_{12}) +$$

$$+ ... + x_m(F_{2m} - F_{1m}) = \sum_{i=1}^{m} x_i V_{1i} < 0,$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \qquad (13)$$

$$\zeta_{n-1}(X) = R_n(X) - R_{n-1}(X) =$$

$$= x_1(F_{n1} - F_{n-1,1}) + ... + x_m(F_{nm} - F_{n-1,m}) =$$

$$= \sum_{i-1}^{m} x_i V_{n-1,} i < 0,$$

где $v_{ji} = F_{j+1,j} - F_{ji,}\ j = 1, 2, ..., n-1, i = 1, 2, ..., m.$

При этом значение $\zeta_j, j = 1, 2, ..., n-1$, характеризует величину разности «полезностей» для $(j+1)$ – го и j – го объектов.

Система неравенств (13) добавлением положительных переменных $x_{m+j}, j = 1, 2, ..., n-1,$ преобразуется в систему равенств

$$\sum_{i=1}^{m} x_i V_{ji} + x_{m+j},\ j = 1, 2, ..., n-1. \qquad (14)$$

Таким образом, задача сведена к поиску решения однородной системы линейных алгебраических уравнений (14). Эта система всегда имеет тривиальное решение:

$$x_1 = x_2 = ... = x_n = x_{n+1} = ... = x_{n+m-1} = 0.$$

Это решение – единственное, если ранг основной матрицы системы равен числу переменных. В противном случае существует бесчисленное множество решений.

Особенность задачи состоит в том, что к искомому решению предъявляется специфическое требование – неотрицательность переменных $x_1, x_2, ..., x_m$ и положительность переменных $x_{m+j}, j = 1, 2, ..., n-1.$

Число дополнительных переменных можно сократить до одного, если в качестве решения системы использовать чебышевскую точку [7].

Теперь, с учетом требований к знакам переменных, решение задачи можно получить с использованием методов линейного программирования: найти набор X, минимизирующий $x_m + 1$ и удовлетворяющий системе линейных уравнений

$$\sum_{i=1}^{m} x_i V_{ji} + x_{m+1} = 0,\ j = 1,2,...,n-1. \qquad (15)$$

При этом, с целью исключения тривиального решения $x_1 = x_2 = ... = x_m = x_{m+1} = 0$ к системе (15) следует добавить ещё одно уравнение

$$\sum_{i=1}^{m} x_i = 1, \qquad (16)$$

являющееся условием нормировки для коэффициентов уравнения регрессии (11), а также неравенство $x_{m+1} > 0$. Получаемое решение используется для расчета «полезности» объектов по формуле (11).

Г. Компараторная идентификация для случая, когда независимые переменные заданы нечетко.

Во многих практических ситуациях значения характеристик объектов не могут быть определены точно. Для описания этих значений используем технологию нечеткой математики [5,8]. Будем считать, что значениеi-й характеристики j-го объекта есть нечеткое число r_{ji} с функцией принадлежности

$$\mu_{ji}(r_{ji}), i = 1,2,...,m, j = 1,2,...,n.$$

Поскольку характеристики объектов – нечеткие числа, то нечеткими является и результаты расчетов значений уровня «полезности» объектов

$$Q_j(x) = \sum_{i=1}^{m} x_i r_{ji},\ i = 1,2,...,m. \qquad (17)$$

Пусть, для определенности, r_{ji} – нечеткие числа с треугольной функцией принадлежности, то есть

$$\mu_{ij}(r_{ij}) = \begin{cases} 0, r_{ji} < a_{ji}, \\ \dfrac{r_{ji} - a_{ji}}{c_{ji} - a_{ji}}, a_{ji} \le r_{ji} < c_{ji}, \\ \dfrac{b_{ji} - r_{ji}}{b_{ji} - c_{ji}}, c_{ji} \le r_{ji} < b_{ji}, \\ 1, r_{ji} \ge b_{ji}. \end{cases} \qquad (18)$$

Поскольку треугольные нечеткие числа – частный случай чисел с функцией принадлежности $(L-R)$- типа, то для приближенных расчетов можно использовать правила выполнения операций, принятых для нечетких чисел этого типа.

С учетом этого, определим функции принадлежности нечетких чисел $Q_i(X)$:

$$\mu_i(Q_i(X)) = \begin{cases} 0, & Q_j(X) < A_i, \\ \dfrac{Q_i(X) - A_i}{C_i - A_i}, & A_i \le Q_i(X) < C_i, \\ \dfrac{B_i - Q_i(X)}{B_i - C_i}, & C_i \le Q_i(X) < B_i, \\ 0, & Q_i(X) \ge B_i \end{cases} \qquad (19)$$

$$A_i = \sum_{j=1}^{n} x_j a_{ji},\ C_i = \sum_{j=1}^{n} x_j c_{ji},\ B_i = \sum_{j=1}^{n} x_j b_{ji}.$$

Если в рассматриваемом случае ранжировка (11) объектов по уровню их «полезности» сохранилась, то естественным аналогом (13) будут нечеткие неравенства

$$\zeta_1(X) = Q_2(X) - Q_1(X) =$$
$$= x_1(r_{21} - r_{11}) + x_2(r_{22} - r_{12}) +$$
$$+ ... + x_m(r_{2m} - r_{1m}) =$$
$$= \sum_{i=1}^{m} x_i w_{ji} \langle 0,$$
$$\dots\dots\dots\dots\dots \qquad (20)$$
$$\zeta_{n-1}(X) = Q_n(X) - Q_{n-1}(X) =$$
$$= \sum_{i=1}^{m} x_i w_{n-1,i} < 0,$$

$$w_{ji} = r_{j+1,i} - r_{ji},$$
$$j = 1,2,...,n-1, i = 1,2,....,m. \qquad (21)$$

Поставим задачу отыскания неотрицательного набора $X = (x_1, x_2, ..., x_n)$, обеспечивающего выполнение неравенств (20).

Система неравенств (20) с добавлением положительной переменной x_{m+1} преобразуется в нечеткую систему линейных алгебраических уравнений

$$\sum_{i=1}^{m} x_i w_{ij} + x_{m+1} = 0, j = 1,2,...,n-1, \qquad (22)$$

где w_{ji}- нечеткие числа, функция принадлежности которых определяется с учетом (22).

Используя стандартное описание функции принадлежности треугольных чисел $L-R$-типа в виде

$$r_{ji} =< c_{ji}, \alpha_{ji}, \beta_{ji} >_{L-R},$$

где c_{ji}-мода числа $r_{ji}, \alpha_{ji} = c_{ji} - a_{ji}, \beta_{ji} = b_{ji} - c_{ji}$, и правила выполнения операций, принятых для нечетких чисел $(L-R)$-типа, запишем функцию принадлежности числа w_{ji}.

Имеем

$$w_{ji} = r_{j,i+1} - r_{ij} =$$
$$=< c_{j,i+1} - c_{ij}, c_{ij} - a_{ij} + b_{j,i+1} - c_{j,i+1},$$
$$b_{ij} - c_{ij} + c_{j,i+1} - a_{j,i+1} >_{LR} =< m_{ji}, \alpha_{ji}, \beta_{ji} >_{LR}$$

Здесь m_{ji}-мода числа w_{ji}, α_{ji} и β_{ji}-левый и правый коэффициент нечеткости.

Введем нечеткие числа

$$Z_j(X) = \sum_{i=1}^{m} x_i w_{ij} + x_{m+1}, j = 1,2,....,n-1,$$

и запишем функции их принадлежности:

$$\mu(Z_j(X)) = \begin{cases} 0, & Z_j(X) < a_1 \\ \dfrac{Z_j(x) - a_1}{\sum\limits_{i=1}^{m} x_i \alpha_{ji}}, & a_1 \le Z_j(X) \le a_2, \\ \dfrac{a_3 - Z_j(X)}{\sum\limits_{i=1}^{m} x_i \beta_{ji}}, & a_2 \le Z_j(x) \le a_3, \\ 0, & Z_i(x) \ge a_3; \end{cases} \quad (23)$$

$$a_1 = \sum_{i=1}^{m} x_i (m_{ji} - \alpha_{ji}) + x_{m+1}, \; a_2 = \sum_{i=1}^{m} x_i m_{ji} + x_{m+1},$$

$$a_3 = \sum_{j=1}^{n} x_j (m_{ji} + \beta_{ji}) + x_{n+1}.$$

Преобразуем вытекающую из (22) систему нечетких уравнений

$$Z_j(X) = 0, j = 1,2,....,n-1,$$

в обычную систему линейных алгебраических уравнений задав, нечеткие числа W_{ji} равными их модальным значениям. При этом получим

$$\sum_{i=1}^{m} x_i m_{ji} + x_{m+1} = 0, j = 1,2,...,n-1. \quad (24)$$

С целью исключения тривиального решения $x_i = 0, i = 1,2,...,m+1$, системы (24) добавим к ней ещё одно нормирующее уравнение (6).

Пусть набор $X^{(0)} = (x_0^{(0)}, x_1^{(0)}, ..., x_m^{(0)}, x_{m+1}^{(0)})$ есть решение системы (24), (16).

Используем введенное в [6] определение «четкое решение нечеткой системы линейных алгебраических уравнений». В соответствии с этим определением четким решением системы уравнений (24), (16) будем называть набор $X = (x_1, x_2, ..., x_{n+1})$, минимизирующий сумму площадей фигур, ограниченных функциями принадлежности $\mu(Z_j)$ нечетких чисел $Z_1, Z_2, ..., Z_{n-1}$, и наименее уклоняющийся от $X^{(0)}$.

Смысл этого определения понятен. Его использование обеспечивает получение набора четких чисел $(x_1, x_2, ..., x_{n+1})$, максимально близкого к модальному $X^{(0)}$, и для которого функции принадлежности нечетких чисел $Z_1, Z_2, ..., Z_{m-1}$ наименее размыты. В качестве критерия компактности функций принадлежности нечетких чисел $Z_j, j = 1,2,...,n-1$, могут быть использованы квадраты длины интервалов – носителей соответствующих нечетких чисел. Тогда мера качества решения системы (24), (16) будет иметь вид:

$$J = \sum_{i=1}^{n} x_i^2 \Big(\sum_{i=1}^{m-1} (\beta_{ij} + \alpha_{ij}) \Big)^2 + \sum_{i=1}^{m} (x_i - x_i^{(0)})^2 \quad (25)$$

Минимизация (25) совместно с условием нормировки (16) дает искомый набор

$$X = (x_1, ..., x_{m+1}).$$

Список литературы

1. *Вучков И. Прикладной линейный регрессионный анализ / Пер. с болг. И. Вучков, Л. Бояджиева, Б. Солаков. – М.: Финансы и статистика, 1987. – 239 с.*
2. *Rao J. Subrahmaniam K. Combining independent estimators and estimation in lincar regression with unegual variances // Biometrics. – 1997. – V. 27. – P. 971*
3. *Durbin J. Errors in variables / J. Durbin // Rev. Int. Stat. Inst. – 1954. – V. 22. – P. 23.*
4. *Madansky A. The bitting of straight lines when both variables are subject to error / A. Madansky // J. Amer. Stat. Assoc. – 1959. – V. 54. – P. 173*
5. *Раскин Л.Г. Нечеткая математика. Основы теории. Приложения / Л.Г. Раскин, О.В. Серая. – Х.: Парус, 2008. – 352 с.*
6. *Ведение в нормативную теорию принятия решений / В.В. Крючковский, Э.Г. Петров, Н.А. Соколова, В.Е. Ходаков. – Херсон : Гринь Д.С., 2013. – 284 с.*
7. *Зуховицкий С.И. Линейное и выпуклое программирование / С.И. Зуховицкий, Л.И. Авдеева. – М.: Наука, 1967. – 460 с.*
8. *Zadeh L.A. Fuzzy sets / Zadeh L.A. // Inf. Conf. – 1965. – 8. – P. 338-353.*

Поступила в редколлегию 25.08.2015

Рецензент: д-р техн. наук, проф. С.М. Порошин, Национальный технический университет «ХПИ» , Харьков.

ІНФОРМАЦІЙНІ ПРОБЛЕМИ КАНОНІЧНОГО РЕГРЕСІЙНОГО АНАЛІЗУ

Л.Г. Раскін, О.В. Сіра

Розглянуто технології вирішення задач регресійного аналізу для випадків, коли порушуються традиційні вихідні передумови . При цьому розроблені методи оцінювання параметрів рівняння регресії , якщо незалежні змінні - випадкові величини або задані нечітко. Запропоновано вирішення завдання компараторної ідентифікації для нечітких вихідних даних .

***Ключові слова**: регресійний аналіз, порушення передумов, нечіткі вихідні дані, компараторного ідентифікація*

INFORMATION PROBLEMS OF CANONICAL REGRESSION ANALYSIS

L.G. Raskin, O.W. Seraja

The technology solutions for the problems of regression analysis where the traditional presuppositions are violated. At the same time methods for estimating the parameters of the regression equation when the independent variables - the randomor fuzzy variables are developed. A solution of the problem of fuzzy comparator for identifying the source data is offered.

***Keywords**: regression analysis, violation of assumptions, fuzzy source data,comparator identification.*