

УДК 517.2+519.2

НОВЫЕ РЕЗУЛЬТАТЫ В АНАЛИЗЕ МНОГОМЕРНЫХ ДАННЫХ: МНОГОМЕРНО-МАТРИЧНЫЙ ПОЛИНОМИАЛЬНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ

В.С. МУХА

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 12 февраля 2019

Аннотация. Обсуждаются новые результаты, полученные в БГУИР в области анализа многомерных данных, применительно к многомерно-матричному полиномиальному регрессионному анализу. Приводится постановка задачи многомерно-матричного полиномиального регрессионного анализа, когда входные и выходные переменные функции регрессии являются многомерными матрицами, и функция регрессии представляется полиномом входной переменной. Приводится также система линейных многомерно-матричных уравнений, решение которой дает оценки параметров полиномиальной функции регрессии. Для многомерно-матричной аффинной функции регрессии получены распределения оценок параметров, позволяющие делать статистические выводы о параметрах функции регрессии и регрессионной модели в целом. Выполнено компьютерное моделирование, иллюстрирующее применение предложенной теории и подтвердившее ее правильность.

Ключевые слова: анализ многомерных данных, многомерные матрицы, регрессионный анализ, дисперсионный анализ, полиномиальная функция регрессии.

Abstract. The new results in the analysis of the multidimensional data received in the BSUIR in the area of the polynomial regression analysis are discussed. The statement of the problem of the regression analysis under conditions of the multidimensional-matrix input and output variables and the polynomial regression function is formulated. The solution to the problem in the form of the system of the linear equations relatively the unknown parameters is given. The properties and distributions of the estimations of the parameters of the affine regression function are investigated, that allow testing hypotheses about the parameters. The computer simulation confirming the theory is performed.

Keywords: analysis of the multidimensional data, multidimensional matrices, regression analysis, dispersion analysis, polynomial regression function.

Doklady BGUIR. 2019, Vol. 121, No. 3, pp. 57-64

**A new results in the analysis of the multidimensional data:
multidimensional-matrix polynomial regression analysis**

V.S. Mukha

Введение

Анализ многомерных данных представляет собой область знаний, в которой исследуются явления, системы и процессы, описываемые наборами переменных. Начало исследований в этой области знаний в БГУИР положено публикацией статьи [1]. Развитие направления можно проследить по работам [2–4]. Особенностью исследований в БГУИР является использование многомерно-матричного математического подхода [2], когда переменные математических моделей данных рассматриваются как многомерные матрицы. Многомерно-матричный подход преодолевает недостатки, присущие классическим скалярному и векторно-матричному подходам, важнейшими из которых являются громоздкость, плохая формализованность при построении нелинейных математических моделей данных

и, как следствие плохой формализованности, отсутствие алгоритмической общности [2]. В статье приводятся новые результаты, полученные в БГУИР в области многомерно-матричного полиномиального регрессионного анализа.

В настоящее время достаточно хорошо развита теория множественной (векторно-скалярной) регрессии [5]. Однако эта теория сталкивается с проблемой плохой формализованности при необходимости рассмотрения нелинейных по входным переменным функций регрессии. Скалярное представление функции регрессии многих переменных отрезком ряда Тейлора произвольной степени весьма громоздко, векторно-матричное представление отсутствует, а представление функции регрессии в виде скалярного произведения вектора параметров и вектора базисных функций, представляющего собой интерпретацию отрезка ряда Тейлора в скалярной форме, является плохо формализованным, так как отсутствует формула, определяющая вектор базисных функций. В любом из этих случаев не обеспечивается алгоритмическая общность, то есть работоспособность алгоритмов, реализующих теорию в виде компьютерных программ, для любого числа переменных и любых степеней аппроксимирующих полиномов. Исследования в области многомерного, в частности, векторно-векторного регрессионного анализа относятся, в основном, к линейной по параметрам и входным переменным функции регрессии и выглядят гораздо беднее [6]. Оценки параметров для этого случая в литературе известны, однако процедуры проверки гипотез отсутствуют.

Многомерно-матричный подход позволяет разрабатывать теорию регрессионного анализа, свободную от указанных недостатков, и в более общей форме, когда входные и выходные переменные функции регрессии являются многомерно-матричными. Такая форма представления функции регрессии включает классические множественную и многомерную регрессии. В работе [7] впервые дана постановка задачи многомерно-матричного полиномиального регрессионного анализа, получены система уравнений для оценок параметров функции регрессии и выражения для оценок параметров постоянной, аффинной и квадратичной регрессий. В работе [8] найдены распределения оценок параметров аффинной функции регрессии. Данная статья содержит полученные ранее результаты без их доказательств, не опубликованные ранее процедуру дисперсионного анализа для аффинной регрессии и результаты компьютерного моделирования. Это позволяет получить более полное представление о проблеме и степени ее разработки.

Постановка и решение задачи многомерно-матричного полиномиального регрессионного анализа

Пусть $x = (x_j)$, $j = (j_1, j_2, \dots, j_q)$, $j_1 = \overline{1, m_1}$, ..., $j_q = \overline{1, m_q}$, – q -мерно-матричная входная переменная некоторого объекта, $\eta = (\eta_i)$, $i = (i_1, i_2, \dots, i_p)$, $i_1 = \overline{1, n_1}$, ..., $i_p = \overline{1, n_p}$, – p -мерно-матричная выходная переменная этого же объекта, и выходная переменная η имеет стохастическую зависимость от входной переменной x так, что существует неизвестная условная плотность вероятности $f(\eta|x)$. Функцию регрессии η на x обозначим $y = \varphi(x)$ и предположим, что плотность вероятности $f(\eta|x)$ можно представить в виде $\eta = \varphi(x) + \xi$, где $\xi = (\xi_i)$, $i = (i_1, i_2, \dots, i_p)$ – p -мерная случайная матрица с нулевым математическим ожиданием. Пусть для некоторых значений $x_1 = (x_{j,1})$, $x_2 = (x_{j,2})$, ..., $x_n = (x_{j,n})$ входной переменной x получены значения $y_{o,1} = (y_{o,i,1})$, $y_{o,2} = (y_{o,i,2})$, ..., $y_{o,n} = (y_{o,i,n})$ выходной переменной η в виде

$$y_{o,\mu} = \varphi(x_\mu) + z_\mu, \quad \mu = \overline{1, n}, \quad (1)$$

где $z_\mu = (z_{i,\mu})$ – значения случайной матрицы $\xi = (\xi_i)$ (ошибки измерений), и по измерениям $(x_1, y_{o,1}), (x_2, y_{o,2}), \dots, (x_n, y_{o,n})$ нужно получить математическую модель объекта в виде эмпирической функции регрессии $\hat{y} = \hat{\varphi}(x)$.

Пусть гипотетическая функция регрессии $y = \varphi(x)$ является полиномом степени m переменной $\tilde{x}^k = x^k - s_{x^k}$, то есть

$$y = \varphi(x) = \sum_{k=0}^m {}^{0,kq} (C_{(p,kq)} \tilde{x}^k) = \sum_{k=0}^m {}^{0,kq} (\tilde{x}^k C_{(kq,p)}), \quad m = 0, 1, 2, \dots, \quad (2)$$

где $C_{(p,kq)}$ – $(p+kq)$ -мерные матрицы коэффициентов, $C_{(p,kq)} = (c_{i,j_1, \dots, j_k}) = (c_{i, \bar{j}_k})$, $i = (i_1, i_2, \dots, i_p)$, $\bar{j}_k = (j_1, j_2, \dots, j_k)$, симметричные относительно q -мультииндексов j_1, j_2, \dots, j_k , и $\tilde{x}^k = x^k - s_{x^k}$, $s_{x^k} = \frac{1}{n} \sum_{\mu=1}^n x_{\mu}^k$. Матрицы $C_{(p,kq)}$ и $C_{(kq,p)}$ должны удовлетворять условиям $C_{(p,kq)} = (C_{(kq,p)})^{H_{p+kq,kq}}$, $C_{(kq,p)} = (C_{(p,kq)})^{B_{p+kq,kq}}$, где $H_{p+kq,kq}$ и $B_{p+kq,kq}$ – подстановки транспонирования типа «назад» и «вперед» соответственно [4].

Подставляя (2) в (1), получим математическую модель измерений в виде

$$y_{o,\mu} = \sum_{k=0}^m {}^{0,kq} (C_{(p,kq)} \tilde{x}_{\mu}^k) + z_{\mu} = \sum_{k=0}^m {}^{0,kq} (\tilde{x}_{\mu}^k C_{(kq,p)}) + z_{\mu}, \quad \mu = \overline{0, n}, \quad m = 0, 1, 2, \dots \quad (3)$$

В предположениях модели измерений (3) требуется по измерениям $(x_1, y_{o,1})$, $(x_2, y_{o,2})$, ..., $(x_n, y_{o,n})$ получить оценки $\hat{C}_{(p,0)}$, $\hat{C}_{(p,q)}$, ..., $\hat{C}_{(p,mq)}$ неизвестных параметров $C_{(p,0)}$, $C_{(p,q)}$, ...,

$C_{(p,mq)}$ методом наименьших квадратов: $f = \sum_{\mu=1}^n {}^{0,p} (z_{\mu} z_{\mu}) \rightarrow \min_{C_{(p,0)}, C_{(p,q)}, \dots, C_{(p,mq)}}$, где

$$z_{\mu} = (z_{i,\mu}) = y_{o,\mu} - \sum_{k=0}^m {}^{0,kq} (C_{(p,kq)} x_{\mu}^k) = y_{o,\mu} - \sum_{k=0}^m {}^{0,kq} (x_{\mu}^k C_{(kq,p)}); \quad \mu = \overline{1, n}, \quad i = (i_1, i_2, \dots, i_p).$$

Решение сформулированной задачи следующее [8]: оценка $\hat{C}_{(p,0q)}$ параметра $C_{(p,0q)}$ модели измерений (3) имеет вид

$$\hat{C}_{(p,0q)} = s_y = \frac{1}{n} \sum_{\mu=1}^n y_{o,\mu}, \quad (4)$$

а оценки $\hat{C}_{(p,kq)}$ параметров $C_{(p,kq)}$ при $k > 0$ определяются из системы уравнений

$$\sum_{k=1}^m {}^{0,kq} (C_{(p,kq)} s_{\tilde{x}^{k+\lambda}}) = s_{y \tilde{x}^{\lambda}}, \quad \lambda = \overline{1, m}, \quad (5)$$

где $s_{y \tilde{x}^{\lambda}} = \frac{1}{n} \sum_{\mu=1}^n y_{o,\mu} x_{\mu}^{\lambda}$, $\lambda = \overline{0, m}$; $s_{x^{k+\lambda}} = \frac{1}{n} \sum_{\mu=1}^n x_{\mu}^k x_{\mu}^{\lambda} = \frac{1}{n} \sum_{\mu=1}^n x_{\mu}^{k+\lambda}$, $s_{x^{\lambda}} = \frac{1}{n} \sum_{\mu=1}^n x_{\mu}^{\lambda}$, $\lambda = \overline{0, m}$.

В статье [7] на основании системы уравнений (5) получены выражения для оценок параметров постоянной, аффинной и квадратичной функций регрессии.

Аффинная функция регрессии, распределения оценок параметров

В статье [8] выполнен более детальный анализ аффинной многомерно-матричной функции регрессии

$$y = C_{(p,0q)} + {}^{0,q} (C_{(p,q)} \tilde{x}_{\mu}) = C_{(0q,p)} + {}^{0,q} (\tilde{x}_{\mu} C_{(q,p)}). \quad (6)$$

Модель измерений в этом случае выглядит следующим образом:

$$y_{o,\mu} = C_{(p,0q)} + {}^{0,q} (C_{(p,q)} \tilde{x}_{\mu}) + z_{\mu} = C_{(0q,p)} + {}^{0,q} (\tilde{x}_{\mu} C_{(q,p)}) + z_{\mu}, \quad \mu = \overline{1, n}, \quad (7)$$

Оценка $\hat{C}_{(p,q)}$ параметра $C_{(p,q)}$ функции регрессии (6) определяется выражением

$$\hat{C}_{(p,q)} = {}^{0,q} (s_{y \tilde{x}} (s_{\tilde{x}^2})^{-1}), \quad (8)$$

где $s_{y\bar{x}} = \frac{1}{n} \sum_{\mu=1}^n y_{o,\mu} \bar{x}_\mu$, $s_{\bar{x}^2} = s_{\bar{x}^{1+1}} = \frac{1}{n} \sum_{\mu=1}^n \bar{x}_\mu^2$, $(s_{\bar{x}^2})^{-1}$ – матрица, $(0, q)$ -обратная к матрице $s_{\bar{x}^2}$,

а оценка $\hat{C}_{(p,0q)}$ параметра $C_{(p,0q)}$ имеет вид (4).

Относительно оценок параметров аффинной функции регрессии (6) справедлива следующая теорема.

Теорема 1. Если матрицы ошибок измерений z_μ в модели измерений (7) имеют математическое ожидание $E(z_\mu) = 0$, дисперсионную матрицу $D(z_\mu) = D(\xi) = \sigma^2 E(0, p)$, где $E(0, p)$ – $(0, p)$ -единичная $(n_1 \times \dots \times n_p \times n_1 \times \dots \times n_p)$ -матрица, и независимы по μ , то математические ожидания оценок $\hat{C}_{(p,q)}$ и $\hat{C}_{(p,0q)}$ определяются выражениями:

$$E(\hat{C}_{(p,q)}) = C_{(p,q)}; \quad (9)$$

$$E(\hat{C}_{(p,0q)}) = C_{(p,0q)}, \quad (10)$$

а их ковариационные и дисперсионные матрицы – выражениями:

$$\text{cov}(\hat{C}_{(q,p)}, \hat{C}_{(p,q)}) = \frac{\sigma^2}{n} D^{(2)}, \quad D^{(2)} = (d_{j,i,i',j'}^{(2)}) = \begin{pmatrix} (s_{\bar{x}^2})_{j,j'}^{-1}, & i = i'; \\ 0 & \text{иначе} \end{pmatrix},$$

$$D(\hat{C}_{(p,q)}) = \frac{\sigma^2}{n} D^{(1)}, \quad D^{(1)} = (d_{i,j,i',j'}^{(1)}) = (D^{(2)})^{(H_{p+q,q} E_{p+q})}, \quad (11)$$

$$D(\hat{C}_{(p,0q)}) = \text{cov}(\hat{C}_{(p,0q)}, \hat{C}_{(p,0q)}) = D^{(0)} = (d_{i,i'}^{(0)}) = \frac{\sigma^2}{n} E(0, p), \quad (12)$$

$\text{cov}(\hat{C}_{(0q,p)}, \hat{C}_{(p,q)}) = 0$, $(s_{\bar{x}^2})_{j,j'}^{-1}$ – элемент матрицы $(s_{\bar{x}^2})^{-1}$, $i = (i_1, i_2, \dots, i_p)$, $i' = (i'_1, i'_2, \dots, i'_p)$, $j = (j_1, j_2, \dots, j_q)$, $j' = (j'_1, j'_2, \dots, j'_q)$. Если ошибки измерений z_μ распределены по нормальному закону, то оценки $\hat{C}_{(p,q)}$, $\hat{C}_{(p,0q)}$ также распределены по нормальному закону с математическими ожиданиями (9), (10) и дисперсионными матрицами (11), (12) соответственно.

В регрессионном анализе дисперсия ошибок измерений σ^2 обычно неизвестна, и требуется по измерениям получить ее оценку $\hat{\sigma}^2$. Оценкой параметра σ^2 может служить статистика

$$\hat{\sigma}^2 = \frac{1}{n \cdot \text{tr}E(0, p)} \text{tr} \left(\sum_{\mu=1}^n (y_{o,\mu} - \hat{y}_\mu)^2 \right), \quad (13)$$

где $\text{tr}E(0, p) = \prod_{i=1}^p n_i$ – след матрицы $E(0, p)$.

Теорема 2. В условиях теоремы 1 (кроме условия нормальности ошибок измерений) математическое ожидание оценки $\hat{\sigma}^2$ (13) определяется выражением $E(\hat{\sigma}^2) = \sigma^2 - \frac{\sigma^2(1 + \text{tr}E(0, q))}{n}$, то есть оценка $\hat{\sigma}^2$ (13) асимптотически несмещенная.

$$\text{Исправленная оценка } \hat{\sigma}_1^2 = \frac{n\hat{\sigma}^2}{n-1 - \text{tr}(E(0, q))} = \frac{1}{\text{tr}E(0, p)(n-1 - \text{tr}E(0, q))} \text{tr} \left(\sum_{\mu=1}^n {}^{0,0} (y_{o,\mu} - \hat{y}_\mu)^2 \right)$$

параметра σ^2 является несмещенной ($E(\hat{\sigma}_1^2) = \sigma^2$). Если ошибки измерений z_μ имеют нормальное распределение, то статистика

$$v = \frac{n \text{tr}E(0, p) \hat{\sigma}^2}{\sigma^2} = \frac{n \text{tr}E(0, p)(n-1 - \text{tr}E(0, q)) \hat{\sigma}_1^2}{\sigma^2} \quad (14)$$

распределена по закону хи-квадрат с $trE(0, p)(n-1-trE(0, q))$ степенями свободы, где $trE(0, p) = \prod_{i=1}^p n_i$, $trE(0, q) = \prod_{j=1}^q m_j$. Оценки $\hat{\sigma}^2$ (13), $\hat{C}_{(p,0q)}$ (4) и $\hat{C}_{(p,q)}$ (8) попарно независимы.

В [8] рассмотрены также некоторые статистики, связанные с полученными оценками параметров. Для возможности рассмотрения отдельных элементов матриц оценок вводятся более простые обозначения: $C^{(0)} = (c_i^{(0)}) = C_{(p,0q)}$, $\bar{C}^{(0)} = (\bar{c}_i^{(0)}) = \bar{C}_{(p,0q)}$, $C^{(1)} = (c_{i,j}^{(1)}) = C_{(p,q)}$, $\bar{C}^{(1)} = (\bar{c}_{i,j}^{(1)}) = \bar{C}_{(p,q)}$, $i = (i_1, i_2, \dots, i_p)$, $j = (j_1, j_2, \dots, j_q)$. Статистики

$$u_i^{(0)} = \frac{\bar{c}_i^{(0)} - c_i^{(0)}}{\sigma} \sqrt{n}, \quad u_{i,j}^{(1)} = \frac{\bar{c}_{i,j}^{(1)} - c_{i,j}^{(1)}}{\sigma \sqrt{d_{i,j,i,j}^{(1)}}} \sqrt{n}, \quad i = (i_1, i_2, \dots, i_p), \quad j = (j_1, j_2, \dots, j_q), \quad (15)$$

распределены по нормальному закону $N(0,1)$, а статистики

$$t_i^{(0)} = \frac{\bar{c}_i^{(0)} - c_i^{(0)}}{\hat{\sigma}_1} \sqrt{n}, \quad t_{i,j}^{(1)} = \frac{\bar{c}_{i,j}^{(1)} - c_{i,j}^{(1)}}{\hat{\sigma}_1 \sqrt{d_{i,j,i,j}^{(1)}}} \sqrt{n}, \quad i = (i_1, i_2, \dots, i_p), \quad j = (j_1, j_2, \dots, j_q), \quad (16)$$

имеют распределение Стьюдента с $trE(0, p)(n-1-trE(0, q))$ степенями свободы.

Оценка отклика $\hat{y} = (\hat{y}_i) = \bar{C}_{(p,0q)} + {}^{0,q}(\bar{C}_{(p,q)} \bar{x})$, $i = (i_1, i_2, \dots, i_p)$ имеет математическое ожидание вида $E(\hat{y}) = a^{(y)} = (a_i^{(y)}) = C_{(p,0q)} + {}^{0,q}(C_{(p,q)} \bar{x})$, $i = (i_1, i_2, \dots, i_p)$, то есть является несмещенной. Оценка \hat{y} распределена по нормальному закону со средним значением

$$a^{(y)} = (a_i^{(y)}) \quad \text{и} \quad \text{дисперсионной} \quad \text{матрицей} \quad D(\hat{y}) = \frac{\sigma^2}{n} D^{(y)} = \frac{\sigma^2}{n} (d_{i,i'}^{(y)}), \quad \text{где} \\ D^{(y)} = (d_{i,i'}^{(y)}) = E(0, p) + {}^{0,q}(\bar{x} {}^{0,q}(D^{(2)} \bar{x})).$$

Нормированные статистики

$$u_i^{(y)} = \frac{\hat{y}_i - a_i^{(y)}}{\sigma \sqrt{d_{i,i}^{(y)}}} \sqrt{n}, \quad i = (i_1, i_2, \dots, i_p), \quad (17)$$

распределены по нормальному закону $N(0,1)$. Статистики

$$t_i^{(y)} = \frac{\hat{y}_i - a_i^{(y)}}{\hat{\sigma}_1 \sqrt{d_{i,i}^{(y)}}} \sqrt{n}, \quad i = (i_1, i_2, \dots, i_p), \quad (18)$$

имеют распределение Стьюдента с $trE(0, p)(n-1-trE(0, q))$ степенями свободы.

Выполнив дополнительно k измерений $y_{\delta,1}, y_{\delta,2}, \dots, y_{\delta,k}$ в некоторой точке x , можно получить

$$\text{независимую оценку } \hat{\sigma}_{\delta,0}^2 \text{ параметра } \sigma^2 \text{ по формуле } \hat{\sigma}_{\delta,0}^2 = \frac{1}{trE(0, p)(k-1)} tr \left(\sum_{\mu=1}^k {}^{0,0}(y_{\delta,\mu} - \hat{y}_\delta)^2 \right),$$

где $\hat{y}_\delta = \frac{1}{k} \sum_{\mu=1}^k y_{\delta,\mu}$. Оценка $\hat{\sigma}_{\delta,0}^2$ является несмещенной, а статистика

$$w = \frac{(k-1)trE(0, p)\hat{\sigma}_{\delta,0}^2}{\sigma^2} \quad (19)$$

распределена по закону хи-квадрат с $(k-1)trE(0, p)$ степенями свободы. В силу независимости статистик v и w статистика

$$F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_{\delta,0}^2} \quad (20)$$

имеет распределение Фишера с $trE(0, p)(n-1-trE(0, q))$, $trE(0, p)(k-1)$ степенями свободы.

Статистики (14), (15), (16)–(19) и их распределения используются для построения доверительных интервалов и проверки гипотез для параметров многомерно-матричной аффинной эмпирической регрессии, а статистика (20) – для проверки гипотезы об адекватности математической модели, подобно тому, как это выполняется в классическом регрессионном анализе [5, 9].

Дисперсионный анализ для аффинной функции регрессии

Приведем также не опубликованную ранее процедуру дисперсионного анализа для проверки гипотезы о том, что параметр $C_{(p,q)}$ аффинной функции регрессии (6) равен нулю, то есть гипотезу вида $\{H_0, H_1\}$, где

$$H_0 : C_{(p,q)} = 0, \quad (21)$$

а альтернатива $H_1 : C_{(p,q)} \neq 0$ состоит в том, что хотя бы один из элементов матрицы $C_{(p,q)}$ не равен нулю. Для проверки этой гипотезы применяется метод дисперсионного анализа, основанный на анализе так называемых сумм квадратов R_0^2 , R_1^2 , R^2 :

$$R_0^2 = tr\left(\sum_{\mu=1}^n (y_{o,\mu} - \hat{y}_\mu)^2\right) = tr\bar{R}_0^2, \quad \bar{R}_0^2 = \sum_{\mu=1}^n (y_{o,\mu} - \hat{y}_\mu)^2, \quad \text{где } y_{o,\mu} = C_{(p,0q)} + {}^{0,q}(C_{(p,q)}\tilde{x}_\mu) + z_\mu,$$

$$\hat{y}_\mu = \hat{C}_{(p,0q)} + {}^{0,q}(\hat{C}_{(p,q)}\tilde{x}_\mu); \quad R_1^2 = tr\left(\sum_{\mu=1}^n (y_{o,\mu} - s_y)^2\right) = tr\left(\sum_{\mu=1}^n \tilde{y}_{o,\mu}^2\right) = tr\bar{R}_1^2, \quad \bar{R}_1^2 = \sum_{\mu=1}^n (y_{o,\mu} - s_y)^2;$$

$$R^2 = tr\left(\sum_{\eta=1}^n (\hat{y}_\mu - s_y)^2\right) = tr\bar{R}^2, \quad \bar{R}^2 = \sum_{\eta=1}^n (\hat{y}_\mu - s_y)^2.$$

Справедлива следующая теорема.

Теорема 3. Для сумм квадратов R_1^2 , R_0^2 и R^2 выполняется равенство $R_1^2 = R_0^2 + R^2$. В условиях теоремы 1 (с условием нормальности ошибок измерений) случайная величина R_0^2 / σ^2 имеет распределение хи-квадрат с $trE(0, p)(n-1-trE(0, q))$ степенями свободы. При выполнении гипотезы $H_0 : C_{(p,q)} = 0$ случайные величины R_0^2 / σ^2 , R_1^2 / σ^2 и $R^2 / \sigma^2 = (R_1^2 - R_0^2) / \sigma^2$ имеют распределения хи-квадрат с $trE(0, p)(n-1-trE(0, q))$, $trE(0, p)(n-1)$ и $trE(0, p)trE(0, q)$ степенями свободы соответственно и независимы. Отношение $F = \frac{R_1^2 - R_0^2}{trE(0, p)trE(0, q)} / \frac{R_0^2}{trE(0, p)(n-1-trE(0, q))}$ имеет f -распределение (Фишера) с $trE(0, p)trE(0, q)$, $trE(0, p)(n-1-trE(0, q))$ степенями свободы.

Для проверки гипотезы (21) используется стандартная процедура проверки гипотезы на основе правостороннего критерия значимости.

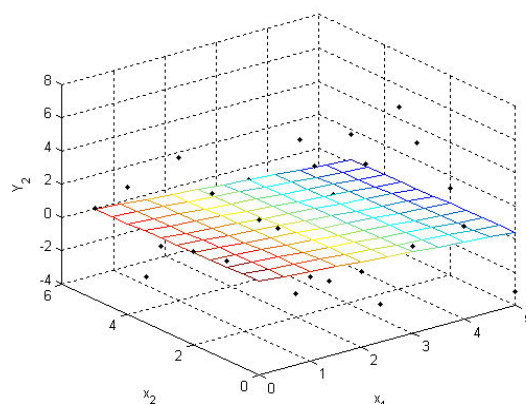
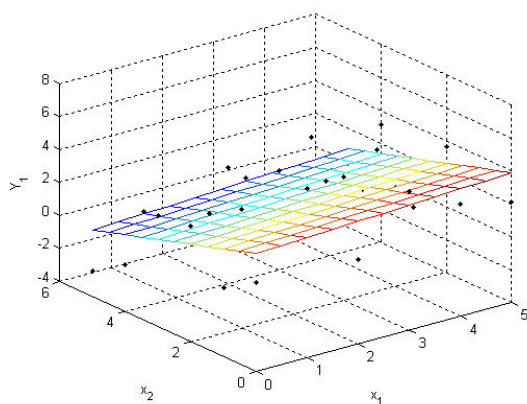
Компьютерное моделирование

Для проверки предложенных алгоритмов моделировалась аффинная функция регрессии (6) с $p = q = 1$, то есть с векторными (в векторно-матричной терминологии) входной и выходной переменными $x = (x_1, x_2)$, $y = (y_1, y_2)$. Математическая модель измерений (7) в этом случае имеет вид

$$y_{o,\mu} = \alpha + {}^{0,1}(\beta x) + z_\mu, \quad \mu = \overline{1, n}, \quad (22)$$

где $z_\mu = (z_{1,\mu}, z_{2,\mu})$. Измерения моделировались на сетке значений входных переменных $0 \leq x_1 \leq 5$, $0 \leq x_2 \leq 5$ с шагом 1 по каждой переменной, что дает выборку размером $n = 36$. Графики эмпирических функций регрессии $y_1 = y_1(x_1, x_2)$, $y_2 = y_2(x_1, x_2)$ при $\alpha = (1, 2)$,

$\beta = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, $\sigma^2 = 5$ представлены на рисунке, на котором также в виде точек изображены измерения, по которым получены данные функции регрессии.



Эмпирические функции регрессии для аффинной многомерной регрессии (22)

Как видно из рисунка, эмпирические функции регрессии не являются горизонтальными плоскостями, то есть визуально мы можем сделать вывод, что существует линейная связь между векторами $y = (y_1, y_2)$ и $x = (x_1, x_2)$. Однако применение изложенных выше процедур проверки гипотез, включая дисперсионный анализ, показывает, что такой связи нет, она признается незначимой. Это соответствует исходным данным, так как матричный параметр β , определяющий линейную связь в истинной функции регрессии, является нулевым. Таким образом, результаты моделирования подтверждают правильность изложенной теории.

Заключение

В статье обобщены результаты, полученные в БГУИР в регрессионном анализе. Рассмотрены неудовлетворенности, существующие в настоящее время в регрессионном анализе применительно к многомерным данным: громоздкость, плохая формализованность, отсутствие алгоритмической общности. По мнению автора, имеющиеся неудовлетворенности были устранены с помощью многомерно-матричного математического подхода. Полученные теоретические результаты и их компьютерные программные реализации не имеют ограничений на мерность и размеры данных, так что единственным ограничением для их использования могут быть лишь вычислительные мощности, которые, однако, растут в настоящее время с огромной скоростью.

Список литературы

1. Муха В.С. Многомерно-матричные производные и разложение функции нескольких переменных в ряд Тейлора // Автоматика и вычислительная техника. 1987. Вып. 16. С. 65–71.
2. Муха В.С. Анализ многомерных данных. Минск: УП «Технопринт», 2004. 368 с.
3. Муха В.С. Анализ многомерных данных: проблемы, состояние, перспективы // Докл. БГУИР. 2004. № 1 (5). С. 38–49.
4. Муха В.С. Математические модели многомерных данных // Докл. БГУИР. 2014. № 2 (80). С. 143–158.
5. Вучков И.Н., Бояджиева Л., Солаков Е. Прикладной линейный регрессионный анализ. М.: Финансы и статистика, 1987. 238 с.
6. Зельнер А. Байесовские методы в эконометрии. М.: Статистика, 1980. 438 с.
7. Муха В.С. Многомерно-матричный полиномиальный регрессионный анализ. Оценки параметров // Весці НАН Беларусі. 2007. № 1. С. 45–51.
8. Муха В.С. Многомерно-матричный линейный регрессионный анализ: распределения и свойства оценок параметров // Весці НАН Беларусі. Сер. фіз.-мат. навук. 2014. № 2. С. 71–81.
9. Муха В.С. Статистические методы обработки данных: учеб. пособие. Минск: Изд. центр БГУ, 2009. 183 с.

References

1. Muha V.S. Mnogomerno-matrichnye proizvodnye i razlozhenie funktsii neskol'kih peremennykh v rjad Tejlora // Avtomatika i vychislitel'naja tehnika. 1987. Vyp. 16. S. 65–71. (in Russ.)
2. Muha V.S. Analiz mnogomernykh dannykh. Minsk: UP «Tehnoprint», 2004. 368 s. (in Russ.)
3. Muha V.S. Analiz mnogomernykh dannykh: problemy, sostojanie, perspektivy // Dokl. BGUIR. 2004. № 1 (5). S. 38–49. (in Russ.)
4. Muha V.S. Matematicheskie modeli mnogomernykh dannykh // Dokl. BGUIR. 2014. № 2 (80). S. 143–158. (in Russ.)
5. Vuchkov I.N., Bojadzhieva L., Solakov E. Prikladnoj linejnyj regressionnyj analiz. M.: Finansy i statistika, 1987. 238 s. (in Russ.)
6. Zel'ner A. Bajesovskie metody v jekonometrii. M.: Statistika, 1980. 438 s. (in Russ.)
7. Muha V.S. Mnogomerno-matrichnyj polinomial'nyj regressionnyj analiz. Ocenki parametrov // Vesci NAN Belarusi. 2007. № 1. S. 45–51. (in Russ.)
8. Muha V.S. Mnogomerno-matrichnyj linejnyj regressionnyj analiz: raspredelenija i svoystva ocenok parametrov // Vesci NAN Belarusi. Ser. fiz.-mat. navuk. 2014. № 2. S. 71–81. (in Russ.)
9. Muha V.S. Statisticheskie metody obrabotki dannykh: ucheb. posobie. Minsk: Izd. centr BGU, 2009. 183 s. (in Russ.)

Сведения об авторе

Муха В.С., д.т.н., профессор, профессор кафедры информационных технологий автоматизированных систем Белорусского государственного университета информатики и радиоэлектроники.

Information about the author

Mukha V.S., D.Sci, professor, professor of department of automated data processing systems of Belarusian state university of informatics and radioelectronics.

Адрес для корреспонденции

220013, Республика Беларусь,
г. Минск, ул. П. Бровки, 6
Белорусский государственный университет
информатики и радиоэлектроники
тел. +375-17-293-88-23;
e-mail: mukha@bsuir.by
Муха Владимир Степанович

Address for correspondence

220013, Republic of Belarus,
Minsk, P. Brovka st., 6
Belarusian state university
of informatics and radioelectronics
tel. +375-17-293-88-23;
e-mail: mukha@bsuir.by
Mukha Vladimir Stepanovich