Сибирский математический журнал
Январь—февраль, 2010. Том 51, № 1

УДК 519.233.5+519.654

О СВЯЗИ МЕЖДУ КОЭФФИЦИЕНТАМИ ПРОСТОЙ И МНОЖЕСТВЕННОЙ РЕГРЕССИОННЫХ МОДЕЛЕЙ

В. Г. Панов, А. Н. Вараксин

Аннотация. Рассмотрена задача нахождения зависимости между коэффициентами множественной регрессионной модели и коэффициентами простых корреляционных моделей, описывающих зависимость между предикторными переменными. Показано, что искомая зависимость описывается матричным равенством.

Ключевые слова: линейная регрессия, простая регрессия, коэффициенты регрессии, метод наименьших квадратов.

Исследование зависимости переменной-отклика в условиях многофакторного воздействия является актуальной и трудной задачей математической статистики. При этом, как правило, получение точной зависимости переменной-отклика от переменных-факторов (называемых также *предикторами*) оказывается слишком сложным и практически бесполезным даже в случае функциональной зависимости. Кроме того, отсутствие функциональной зависимости между переменными является типичным, если эти переменные имеют вероятностный характер. В этом случае описание ожидаемой зависимости носит приближенный характер и выражается некоторым функциональным уравнением между средними величинами.

Точная постановка соответствующей задачи рассматривается в регрессионном анализе. Наиболее простая и разработанная модель регрессионного анализа предполагает линейную статистическую зависимость между стохастическими переменной-откликом Y и переменными-предикторами $X_1, X_2, \ldots, X_n$. Тогда зависимость Y от $X_1, X_2, \ldots, X_n$ можно моделировать линейным регрессионным уравнением, т. е. уравнением множественной линейной регрессии:

$$y = b_0 + \sum_{k=1}^{n} b_k\, x_k. \tag{1}$$

Постоянные $b_0, b_1, \ldots, b_n$ являются решениями вариационной задачи минимизации среднеквадратичного уклонения ($\mathbf{E}$ — оператор математического ожидания):

$$\min_{b_0, b_1, \ldots, b_n} \mathbf{E}\left(Y - b_0 - \sum_{k=1}^{n} b_k X_k\right)^2. \tag{2}$$

Таким образом, задача нахождения уравнения множественной регрессии переменной-отклика на заданный набор переменных-предикторов меняется, как только меняется набор предикторов или сама переменная-отклик. В регрессионном анализе зависимость между коэффициентами регрессионных моделей, построенных по разным системам предикторов, обычно не рассматривается, так как ожидать какой-либо зависимости в этих моделях в общем случае не приходится.

Однако учет статистической зависимости как между предикторами, так и между откликом и частью предикторов делает проблему поиска связи между коэффициентами линейных регрессионных моделей более содержательной.

А именно, рассмотрим следующую задачу. Пусть имеется набор из $n+1$ случайных величин $X_1, X_2, \ldots, X_n, Y$, у которых, вообще говоря, произвольные распределения (предполагается, однако, что среди случайных величин $X_1, X_2, \ldots,$ X_n нет постоянных). В рассматриваемых дальше линейных регрессионных моделях переменные $X_1, X_2, \ldots, X_n$ являются объясняющими (предикторными), а Y — откликом. Для оценки зависимости отклика Y от $X_1, X_2, \ldots, X_n$ рассмотрим линейное регрессионное уравнение, т. е. уравнение множественной линейной регрессии (1).

Будем считать, что для любой пары объясняющих переменных X_i, X_j найдено уравнение линейной регрессии

$$x_i = c_{0ij} + c_{ij}x_j, \quad i, j = 1, 2, \ldots, n, \tag{3}$$

где коэффициенты c_{0ij}, c_{ij} являются решениями вариационной задачи

$$\min_{c_{0ij}, c_{ij}} \mathbf{E}(X_i - c_{0ij} - c_{ij}X_j)^2, \quad i, j = 1, 2, \ldots, n. \tag{4}$$

Допустим также, что известно уравнение простой регрессии Y по каждой из переменных X_k:

$$y = a_{0k} + a_k x_k, \quad k = 1, 2, \ldots, n. \tag{5}$$

Как и раньше, коэффициенты a_{0k}, a_k являются решениями задачи минимизации функционала среднеквадратичного уклонения

$$\min_{a_{0k}, a_k} \mathbf{E}(Y - a_{0k} - a_k X_k)^2, \quad k = 1, 2, \ldots, n. \tag{6}$$

При обычных условиях $[1, 2]$ каждая из этих экстремальных задач имеет единственное решение, причем для задачи (4) выполняются равенства $c_{0ii} = 0$ и $c_{ii} = 1$.

Рассматриваемый в предлагаемой статье вопрос состоит в исследовании связи между коэффициентами введенных выше регрессионных моделей: существует ли какая-либо связь между коэффициентами $\{a_k\}, \{b_k\}, \{c_{ij}\}$?

Предполагаемая связь регрессионных коэффициентов возникла как гипотеза при обработке экспериментальных данных в проблемах медико-биологического мониторинга $[3, 4]$. В качестве исследуемой переменной Y рассматривалась некоторая интегральная характеристика, которую можно назвать заболеваемостью населения, а в качестве факторов $X_1, X_2, \ldots, X_n$ — загрязнение атмосферного воздуха различными токсикантами. В этих работах отмечалось, что в задачах мониторинга модели множественной регрессии должны рассматриваться в условиях согласованного изменения предикторов $\{X_i\}$. В работе [4] предложены два варианта согласованного изменения $\{X_i\}$: пропорциональное и корреляционное. В данной работе изучаются соотношения между коэффициентами при корреляционной зависимости предикторов.

С более формальной точки зрения задача состоит в исследовании зависимости между моделями множественной регрессии отклика на предикторы, парных корреляций предикторов и простых корреляций отклика на каждый из предикторов.

Перед формулировкой основного утверждения докажем вспомогательные леммы. Для случайной величины X обозначим через $\overline{X}$ среднее значение (математическое ожидание) этой случайной величины. Так как будем рассматривать только дискретные случайные величины, ниже во всех случаях среднее значение соответствующей случайной величины существует. Для заданного набора предикторов $X_1, X_2, \dots, X_n$ и отклика Y введем следующую вспомогательную матрицу порядка $n+3$:

$$A = \begin{pmatrix} 1 & \overline{X_1} & \overline{X_2} & \dots & \overline{X_n} & \overline{Y} & 0 \\ \overline{X_1} & \overline{X_1^2} & \overline{X_1X_2} & \dots & \overline{X_1X_n} & \overline{X_1Y} & 0 \\ \overline{X_2} & \overline{X_2X_1} & \overline{X_2^2} & \dots & \overline{X_2X_n} & \overline{X_2Y} & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \overline{X_n} & \overline{X_nX_1} & \overline{X_nX_2} & \dots & \overline{X_n^2} & \overline{X_nY} & 0 \\ 0 & \overline{X_1} & \overline{X_2} & \dots & \overline{X_n} & \overline{Y} & 1 \\ 0 & \overline{X_kX_1} & \overline{X_kX_2} & \dots & \overline{X_kX_n} & \overline{X_kY} & \overline{X_k} \end{pmatrix}. \tag{7}$$

Предпоследняя $(n+2)$-я строка матрицы A получается из ее первой строки перестановкой первого и последнего элементов. Таким же образом из $(k+1)$-й ее строки получается последняя $(n+3)$-я строка.

Лемма 1. *Матрица A вырожденна.*

Доказательство. Раскроем определитель матрицы A по последнему столбцу:

$$\det(A) = (-1)^{2n+5} \begin{vmatrix} 1 & \overline{X_1} & \overline{X_2} & \dots & \overline{X_n} & \overline{Y} \\ \overline{X_1} & \overline{X_1^2} & \overline{X_1X_2} & \dots & \overline{X_1X_n} & \overline{X_1Y} \\ \overline{X_2} & \overline{X_2X_1} & \overline{X_2^2} & \dots & \overline{X_2X_n} & \overline{X_2Y} \\ & & \dots & \dots & & \\ \overline{X_n} & \overline{X_nX_1} & \overline{X_nX_2} & \dots & \overline{X_n^2} & \overline{X_nY} \\ 0 & \overline{X_kX_1} & \overline{X_kX_2} & \dots & \overline{X_kX_n} & \overline{X_kY} \end{vmatrix} + \overline{X_k} \begin{vmatrix} 1 & \overline{X_1} & \overline{X_2} & \dots & \overline{X_n} & \overline{Y} \\ \overline{X_1} & \overline{X_1^2} & \overline{X_1X_2} & \dots & \overline{X_1X_n} & \overline{X_1Y} \\ \overline{X_2} & \overline{X_2X_1} & \overline{X_2^2} & \dots & \overline{X_2X_n} & \overline{X_2Y} \\ & & \dots & \dots & & \\ \overline{X_n} & \overline{X_nX_1} & \overline{X_nX_2} & \dots & \overline{X_n^2} & \overline{X_nY} \\ 0 & \overline{X_1} & \overline{X_2} & \dots & \overline{X_n} & \overline{Y} \end{vmatrix}.$$

В первом определителе вычитаем последнюю строку из $(k+1)$-й строки, а во втором вычитаем последнюю строку из первой строки:

$$\det(A) = -\begin{vmatrix} 1 & \overline{X_1} & \overline{X_2} & \dots & \overline{X_n} & \overline{Y} \\ \overline{X_1} & \overline{X_1^2} & \overline{X_1X_2} & \dots & \overline{X_1X_n} & \overline{X_1Y} \\ & & \dots & \dots & & \\ \overline{X_k} & 0 & 0 & \dots & 0 & 0 \\ & & \dots & \dots & & \\ 0 & \overline{X_kX_1} & \overline{X_kX_2} & \dots & \overline{X_kX_n} & \overline{X_kY} \end{vmatrix}$$

$$+\overline{X_k}\begin{vmatrix} 1 & \overline{0} & \overline{0} & \dots & \overline{0} & \overline{0} \\ \overline{X_1} & \overline{X_1^2} & \overline{X_1X_2} & \dots & \overline{X_1X_n} & \overline{X_1Y} \\ \overline{X_2} & \overline{X_2X_1} & \overline{X_2^2} & \dots & \overline{X_2X_n} & \overline{X_2Y} \\ & & \dots & \dots & & \\ \overline{X_n} & \overline{X_nX_1} & \overline{X_nX_2} & \dots & \overline{X_n^2} & \overline{X_nY} \\ 0 & \overline{X_1} & \overline{X_2} & \dots & \overline{X_n} & \overline{Y} \end{vmatrix}$$

$$= (-1)^{k+1}\overline{X}_k \begin{vmatrix} \overline{X_1} & \overline{X_2} & \dots & \overline{X_n} & \overline{Y} \\ \overline{X_1^2} & \overline{X_1X_2} & \dots & \overline{X_1X_n} & \overline{X_1Y} \\ & & \dots & & \\ \overline{X_{k-1}X_1} & \overline{X_{k-1}X_2} & \dots & \overline{X_{k-1}X_n} & \overline{X_{k-1}Y} \\ \overline{X_{k+1}X_1} & \overline{X_{k+1}X_2} & \dots & \overline{X_{k+1}X_n} & \overline{X_{k+1}Y} \\ & & \dots & & \\ \overline{X_nX_1} & \overline{X_nX_2} & \dots & \overline{X_n^2} & \overline{X_nY} \\ \overline{X_kX_1} & \overline{X_kX_2} & \dots & \overline{X_kX_n} & \overline{X_kY} \end{vmatrix}$$

$$+\overline{X_k}\begin{vmatrix} \overline{X_1^2} & \overline{X_1X_2} & \dots & \overline{X_1X_n} & \overline{X_1Y} \\ \overline{X_2X_1} & \overline{X_2^2} & \dots & \overline{X_2X_n} & \overline{X_2Y} \\ & & \dots & & \\ \overline{X_nX_1} & \overline{X_nX_2} & \dots & \overline{X_n^2} & \overline{X_nY} \\ \overline{X_1} & \overline{X_2} & \dots & \overline{X_n} & \overline{Y} \end{vmatrix}.$$

В первом определителе последовательными перестановками поднимаем последнюю строку до тех пор, пока она не станет $(k+1)$-й. Во втором определителе последнюю строку поднимаем на место первой. Получаются одинаковые определители с множителями $(-1)^{k+1+n-k} = (-1)^{n+1}$ и $(-1)^n$ соответственно. Лемма доказана.

Для доказательства основной теоремы нам понадобится теорема Лапласа о разложении определителя, которую удобно сформулировать, как в [5] (см. также [6]).

Пусть $Q_{r,n}$ — совокупность всех строго возрастающих последовательностей длины r из $(1,2,\dots,n)$; $A\binom{\alpha}{\beta} = A\binom{\alpha_1,\alpha_2,\dots,\alpha_r}{\beta_1,\beta_2,\dots,\beta_r}$ — определитель подматрицы матрицы A, образованной пересечением строк с номерами из $\alpha = (\alpha_1,\alpha_2,\dots,\alpha_r)$ и столбцов с номерами из $\beta = (\beta_1,\beta_2,\dots,\beta_r)$, где α,β — наборы из $Q_{r,n}$; $A\binom{\alpha^*}{\beta^*}$ — определитель подматрицы матрицы A, стоящей на пересечении строк с номерами из α^* и столбцов с номерами из β^*, где α^*,β^* — наборы из $Q_{n-r,n}$, дополнительные для наборов α и β соответственно: $\alpha\cap\alpha^* = \varnothing$, $\beta\cap\beta^* = \varnothing$, $\alpha\cup\alpha^* = \beta\cup\beta^* = \{1,2,\dots,n\}$; для любого набора $\beta = (\beta_1,\beta_2,\dots,\beta_r)$ обозначим $|\beta| = \sum\limits_{i=1}^{r}\beta_i$.

Теорема Лапласа. *Пусть A — квадратная матрица порядка n, $1\le r\le n$, α — фиксированная последовательность из $Q_{r,n}$. Тогда*

$$\det(A) = (-1)^{|\alpha|}\sum_{\beta\in Q_{r,n}}(-1)^{|\beta|}A\begin{pmatrix}\alpha\\ \beta\end{pmatrix}\cdot A\begin{pmatrix}\alpha^*\\ \beta^*\end{pmatrix}.$$

Возьмем $r = n+1$, $\alpha = (1,2,\dots,n+1)$, $\alpha\in Q_{n+1,n+3}$, и применим к матрице (7) порядка $n+3$ теорему Лапласа. Учитывая лемму 1, получим равенство

$$\sum_{\beta\in Q_{n+1,n+3}}(-1)^{|\beta|}A\begin{pmatrix}1,2,\dots,n+1\\ \beta_1,\beta_2,\dots,\beta_{n+1}\end{pmatrix}\cdot A\begin{pmatrix}n+2,\ n+3\\ *\quad *\end{pmatrix} = 0. \qquad (8)$$

Звездочками в этом равенстве обозначены такие номера (расположенные в порядке возрастания) от 1 до $n+3$, которые вместе с набором β образуют все множество $\{1, 2, \ldots, n+3\}$.

Лемма 2. *Количество наборов β из множества $Q_{n+1,n+3}$, для которых произведение определителей в равенстве* (8) *отлично от нуля, не более $n+1$.*

Доказательство. Если в набор β не входит 1, то набор, дополнительный для β, начинается с 1 и, следовательно,

$$A\begin{pmatrix} n+2 & n+3 \\ 1 & * \end{pmatrix} = 0.$$

С другой стороны, в набор β не может входить число $n+3$, так как в силу строения матрицы A

$$A\begin{pmatrix} 1, 2, \ldots, n+1 \\ \beta_1, \beta_2, \ldots, n+3 \end{pmatrix} = 0.$$

Следовательно, интересующие нас наборы β должны иметь вид

$$(1, \beta_2, \ldots, \beta_n, n+1) \quad \text{или} \quad (1, \beta_2, \ldots, \beta_n, n+2).$$

Так как первый набор определен однозначно и равен $(1, 2, \ldots, n, n+1)$, остается доказать, что наборов второго типа будет ровно n. Из этих наборов выделим случай

$$\beta = (1, 2, \ldots, n, n+2).$$

Для остальных наборов рассматриваемого типа существует единственное число i от 2 до n, (включительно) такое, что

$$\begin{aligned} &\beta_j = j \quad \text{при } j = 1, 2, \ldots, i-1, \\ &\beta_i = i+1 \quad \text{при } j = i, \\ &\beta_j = j+1 \quad \text{при } j = i+1, i+2, \ldots, n+1. \end{aligned} \tag{9}$$

Обозначим через $\mathfrak{b}_i$ последовательность, определяемую условиями (9). Таким образом, наборов второго типа будет ровно столько, сколько значений может принимать индекс i в условиях (9). Следовательно, всего наборов второго типа будет $n-1+1=n$, что и доказывает лемму.

Для формулировки основного результата введем следующие матрицы:

$$\mathscr{A} = (a_1, a_2, \ldots, a_n), \quad \mathscr{B} = (b_1, b_2, \ldots, b_n), \quad \mathscr{C} = (c_{ij})_{n\times n},$$

где $\{a_k\}$,$\{b_k\}$, $\{c_{ij}\}$ — коэффициенты уравнений (1), (3), (5), полученные как решение вариационных задач (2), (4), (6).

Теорема. *Коэффициенты a_k, b_i, c_{ik} связаны равенством*

$$a_k = \sum_{i=1}^{n} b_i \cdot c_{ik} \tag{10}$$

или, в матричных обозначениях,

$$\mathscr{A} = \mathscr{B} \cdot \mathscr{C}.$$

Доказательство. Как известно [1, 7], решение задач (2), (4), (6) сводится к решению следующих систем линейных уравнений:

$$\begin{cases} b_0 + b_1\overline{X_1} + b_2\overline{X_2} + \dots + b_n\overline{X_n} = \overline{Y} \\ b_0\overline{X_1} + b_1\overline{X_1^2} + b_2\overline{X_1X_2} + \dots + b_n\overline{X_1X_n} = \overline{X_1Y} \\ \dots\dots\dots\dots\dots \\ b_0\overline{X_n} + b_1\overline{X_nX_1} + b_2\overline{X_nX_2} + \dots + b_n\overline{X_n^2} = \overline{X_nY}\ , \end{cases} \tag{11}$$

$$\begin{cases} c_{0ij} + c_{ij}\overline{X_j} = \overline{X_i} \\ c_{0ij}\overline{X_j} + c_{ij}\overline{X_j^2} = \overline{X_iX_j}, \end{cases} \quad i, j = 1, 2, \dots, n,$$

$$\begin{cases} a_{0k} + a_k\overline{X_k} = \overline{Y} \\ a_{0k}\overline{X_k} + a_k\overline{X_k^2} = \overline{X_kY}, \end{cases} \quad k = 1, 2, \dots, n.$$

Введем обозначения для следующих определителей:

$$\delta_k = \begin{vmatrix} 1 & \overline{X_k} \\ \overline{X_k} & \overline{X_k^2} \end{vmatrix}, \quad \delta_{1k} = \begin{vmatrix} 1 & \overline{Y} \\ \overline{X_k} & \overline{X_kY} \end{vmatrix}, \quad d_{ij} = d_{ji} = \begin{vmatrix} 1 & \overline{X_i} \\ \overline{X_j} & \overline{X_iX_j} \end{vmatrix},$$

$$\Delta = \begin{vmatrix} 1 & \overline{X_1} & \overline{X_2} & \dots & \overline{X_n} \\ \overline{X_1} & \overline{X_1^2} & \overline{X_1X_2} & \dots & \overline{X_1X_n} \\ & & \dots & & \\ \overline{X_n} & \overline{X_nX_1} & \overline{X_nX_2} & \dots & \overline{X_n^2} \end{vmatrix}.$$

Будем считать, что определители δ_k и Δ не равны нулю, т. е. решения этих систем существуют и единственны. Тогда по формулам Крамера имеют место равенства

$$a_k = \frac{\delta_{1k}}{\delta_k}, \quad c_{ik} = \frac{d_{ik}}{\delta_k}, \quad b_i = \frac{\Delta_i}{\Delta}, \tag{12}$$

где определитель Δ_i получается из определителя Δ заменой $(i+1)$-го столбца столбцом свободных членов из системы (11).

Применим теорему Лапласа к матрице A из леммы 1 и учтем леммы 1 и 2. Как нетрудно подсчитать, для наборов β из леммы 2 имеют место равенства (использованы обозначения из теоремы Лапласа)

$$|(1, 2, \dots, n, n+1)| = \frac{(n+1)(n+2)}{2},$$

$$|(1, 2, \dots, n, n+2)| = \frac{(n+1)(n+2)}{2} + 1,$$

$$|(1, 2, \dots, i-1, i+1, \dots, n+1, n+2)| = \frac{(n+2)(n+3)}{2} - i.$$

Для завершения доказательства рассмотрим определители из равенства (8) и учтем лемму 2. Имеем

$$A\begin{pmatrix} 1, 2, \dots, n+1 \\ 1, 2, \dots, n+1 \end{pmatrix} = \Delta, \quad A\begin{pmatrix} n+2, n+3 \\ n+2, n+3 \end{pmatrix} = -\delta_{1n},$$

$$A\begin{pmatrix} 1, 2, \dots, n+1 \\ 1, 2, \dots, n+2 \end{pmatrix} = \Delta_n, \quad A\begin{pmatrix} n+2, n+3 \\ n+1, n+3 \end{pmatrix} = -d_{n\,k},$$

$$A\begin{pmatrix} 1, 2, \dots, n+1 \\ \mathfrak{b}_i \end{pmatrix} = (-1)^{n-i+1}\Delta_{i-1}, \quad A\begin{pmatrix} n+2, n+3 \\ i, n+3 \end{pmatrix} = -d_{i-1\,k},$$

где $\mathfrak{b}_i$ — набор (9), $i = 2, 3, \dots, n$. Подставляя эти выражения в равенство (8), получаем

$$0 = \sum_{\beta \in Q_{n+1,n+3}} (-1)^{|\beta|} A\begin{pmatrix} 1, 2, \dots, n+1 \\ \beta_1, \beta_2, \dots, \beta_{n+1} \end{pmatrix} \cdot A\begin{pmatrix} n+2, \ n+3 \\ * \quad * \end{pmatrix}$$

$$= (-1)^{\frac{(n+1)(n+2)}{2}} \Delta \cdot (-\delta_{1n}) + (-1)^{\frac{(n+1)(n+2)}{2}+1} \Delta_n \cdot (-d_n)$$

$$+ \sum_{i=2}^{n} (-1)^{\frac{(n+2)(n+3)}{2}-i} (-1)^{n-i+1} \Delta_{i-1} \cdot (-d_{i-1\,k})$$

$$= (-1)^{\frac{(n+1)(n+2)}{2}+1} \left[\Delta \cdot \delta_{1k} - \Delta_n \cdot d_{n\,k} - \sum_{i=2}^{n} (-1)^{n+1-i} (-1)^{n-i+1} \Delta_{i-1} \cdot d_{i-1\,k} \right]$$

$$= (-1)^{\frac{(n+1)(n+2)}{2}+1} \left[\Delta \cdot \delta_{1k} - \Delta_n \cdot d_{n\,k} - \sum_{i=2}^{n} \Delta_{i-1} \cdot d_{i-1\,k} \right].$$

Следовательно,

$$\Delta \cdot \delta_{1k} = \Delta_n \cdot d_{n\,k} + \sum_{i=2}^{n} \Delta_{i-1} \cdot d_{i-1\,k}.$$

Деля это равенство на $\delta_k \cdot \Delta$ и учитывая равенства (12), получим

$$a_k = \sum_{i=1}^{n} b_i \cdot c_{ik},$$

что и требовалось доказать.

Следствие 1. *Если предикторы $X_1, X_2, \dots, X_n$ попарно не коррелированы (т. е. $c_{ik} = 0$ для всех $i \neq k$), то $a_k = b_k$. Таким образом, коэффициенты множественной регрессии в этом случае совпадают с коэффициентами парных корреляций.*

Доказательство следует из того, что в случае нулевых попарных корреляций предикторов матрица $\mathscr{C}$ становится единичной.

Следствие 2. *Если матрица $\mathscr{C}$ обратима, то строку коэффициентов $\mathscr{B} = (b_1, b_2, \dots, b_n)$ уравнения (1) можно выразить через коэффициенты простых корреляций Y по предикторам X равенством*

$$\mathscr{B} = A \cdot \mathscr{C}^{-1}.$$

Свободный член b_0 можно найти, например, из первого уравнения (11).

Замечания. 1. Как показывает следствие 2, с помощью полученного соотношения (10) можно вычислить коэффициенты множественной регрессии отклика Y на множество предикторов $X_1, X_2, \dots, X_n$ через коэффициенты парных корреляций $\mathscr{A}$ и $\mathscr{C}$. В регрессионном анализе существует другой метод сведения уравнения множественной регрессии к серии уравнений парных корреляций, изложенный, например, в [8, разд. 4.1]. Однако в этом методе существенную роль играет вычисление регрессии остатков на остатки, которое в данном случае совсем не требуется. Кроме того, сама процедура вычисления коэффициентов множественной регрессии выполняется последовательно в несколько шагов (тем больше, чем больше предикторов), что приводит к накоплению вычислительной ошибки. В доказанном выражении это достигается с помощью одного матричного произведения.

2. Сам метод из [8] является методом нахождения коэффициентов множественной регрессии и не может быть распространен на вычисление каких-либо других регрессионных коэффициентов (например, парных) по известным коэффициентам множественной регрессии. В то же время равенство $\mathscr{A} = \mathscr{B} \cdot \mathscr{C}$ дает возможность вычисления коэффициентов регрессии отклика на каждый предиктор при известной матрице коэффициентов регрессий между предикторами и коэффициентов множественной регрессии отклика на все предикторы.

ЛИТЕРАТУРА

1. *Магнус Я. Р., Нейдеккер Х.* Матричное дифференциальное исчисление с приложениями к статистике и эконометрике. М.: Физматлит, 2002.

2. *Козлов М. В., Прохоров А. В.* Введение в математическую статистику. М.: Изд-во Моск. ун-та, 1987.

3. *Вараксин А. Н., Маслакова Т. А., Чуканов В. Н., Антонов К. Л.* Регрессионная модель зависимости заболеваемости населения от степени загрязнения атмосферного воздуха // Экологические системы и приборы. 2004. № 4. С. 52–55.

4. *Маслакова Т. А., Вараксин А. Н., Чуканов В. Н.* Интерпретация прогностических регрессионных моделей в области медико-экологического мониторинга // Экологические системы и приборы. 2008. № 2. С. 6–9.

5. *Маркус М., Минк Х.* Обзор по теории матриц и матричных неравенств. М.: Едиториал УРСС, 2004.

6. *Гантмахер Ф. Р.* Теория матриц. М.: Наука, 1988.

7. *Линник Ю. В.* Метод наименьших квадратов и основы математико-статистической теории обработки наблюдений. М.: Наука, 1962.

8. *Дрейпер Н., Смит Г.* Прикладной регрессионный анализ. М.: Финансы и статистика, 1986. Кн. 1.

Статья поступила 7 мая 2008 г., окончательный вариант — 5 декабря 2008 г.

Панов Владимир Григорьевич
Уральский гос. университет, кафедра математического анализа и теории функций,
пр. Ленина, 51, Екатеринбург 620083
`vladimir.panov@usu.ru`

Вараксин Анатолий Николаевич
лаборатория математического моделирования
Института промышленной экологии УрО РАН,
ул. Софьи Ковалевской, 20-а, Екатеринбург 620219, ГСП-594
`varaksin@ecko.uran.ru`