УДК 519.862.6

Аналитические зависимости между коэффициентами детерминации и соотношением дисперсий ошибок исследуемых признаков в модели регрессии Деминга

© М.П. Базилевский

Иркутский государственный университет путей сообщения,
Иркутск, 664074, Россия

Рассмотрена проблема построения регрессионных моделей, в которых все переменные имеют стохастический характер. Для ее решения предложено использовать коэффициент детерминации. Получены аналитические зависимости коэффициентов детерминации от соотношения дисперсий ошибок исследуемых признаков. Поставлена оптимизационная задача, предполагающая максимизацию суммы коэффициентов детерминации каждого уравнения в регрессии Деминга. Дан модельный пример численной обработки регрессии Деминга с ее известными параметрами и ошибками признаков.

Ключевые слова: *модель регрессии Деминга, метод наименьших квадратов, дисперсия ошибок, коэффициент детерминации.*

Введение. Регрессионное моделирование является признанным инструментом обработки статистических данных [1, 2]. Построение регрессионной модели осуществляется в строгой последовательности основных этапов [1, 3]: сбор необходимой статистической информации, определение вида модели (т. е. математической формы связи между ее переменными), оценивание неизвестных параметров и проверка качества модели.

Самым распространенным методом оценивания неизвестных параметров регрессионных моделей является метод наименьших квадратов (МНК) [1–4]. Одной из предпосылок МНК является то, что значения объясняющих переменных, входящих в модель, должны быть неслучайными или детерминированными [3, 4]. Однако на практике не всегда возможно обеспечить выполнение этого условия. Это связано с неточностями, которые могут возникать при регистрации значений как объясняемой, так и объясняющих переменных, даже при использовании современных технических средств [2, 5]. Из этого следует, что значения объясняющих переменных могут иметь стохастическую природу. В этом случае оценки параметров модели, полученные с помощью МНК, могут оказаться смещенными, несостоятельными и неэффективными.

Задача построения регрессионной модели, предполагающей стохастический характер всех входящих в нее переменных, может быть решена с помощью регрессии Деминга. Проблема оценивания таких моделей, хоть и не до конца, но достаточно хорошо изучена [5–13].

Однако следует констатировать, что к настоящему времени практически отсутствуют работы по проблеме оценки качества регрессий Деминга. В данной работе качество таких моделей будем оценивать с помощью традиционного для регрессионного анализа коэффициента детерминации.

Регрессия Деминга. Пусть изучается взаимозависимость между двумя количественными признаками y и x. В результате n опытов или наблюдений получены пары чисел $\left(x_i, y_i\right)$, $i=\overline{1,n}$. Предположим, что оба признака имеют стохастический характер, как результат наличия неконтролируемых ошибок в данных. Это означает, что фактически вместо пар $\left(x_i, y_i\right)$ фиксируются значения

$$x_i = x_i^* + \varepsilon_{x_i},\ i=\overline{1,n}, \tag{1}$$

$$y_i = y_i^* + \varepsilon_{y_i},\ i=\overline{1,n}, \tag{2}$$

где $\left(x_i^*, y_i^*\right)$ — «истинные» значения переменных; ошибки ε_x и ε_y — нормально распределенные случайные величины с нулевыми математическими ожиданиями и постоянными дисперсиями, т. е. $\varepsilon_x \sim N\left(0,\sigma_{\varepsilon_x}^2\right)$ и $\varepsilon_y \sim N\left(0,\sigma_{\varepsilon_y}^2\right)$.

Очевидно, что для получения корректных результатов требуется рассматривать зависимость между «истинными» значениями переменных y^* и x^*. Предположим, что такая связь между переменными задается простейшей линейной зависимостью

$$y_i^* = a + bx_i^*,\ i=\overline{1,n}, \tag{3}$$

где a и b — неизвестные параметры.

Используя соотношения (1)–(2) для модели (3), перейдем к регрессии Деминга

$$\begin{cases} y_i = a + bx_i^* + \varepsilon_{y_i}, \\ x_i = x_i^* + \varepsilon_{x_i}, \end{cases} \tag{4}$$

где x_i^* — неизвестные «истинные» значения объясняющей переменной.

Для оценивания неизвестных параметров модели (4) требуется минимизировать сумму квадратов ошибок ε_x и ε_y. Но поскольку эти ошибки имеют разные дисперсии $\sigma_{\varepsilon_x}^2$ и $\sigma_{\varepsilon_y}^2$, наша оптимизационная задача будет иметь следующий функционал:

$$S = \frac{\sum_{i=1}^{n} \varepsilon_{x_i}^2}{\sigma_{\varepsilon_x}^2} + \frac{\sum_{i=1}^{n} \varepsilon_{y_i}^2}{\sigma_{\varepsilon_y}^2} \to \min. \qquad (5)$$

Введем параметр $\lambda = \sigma_{\varepsilon_y}^2 / \sigma_{\varepsilon_x}^2$. Тогда, используя ошибки модели (4), представим функционал (5) в виде

$$S = \frac{1}{\sigma^2}\left(\sum_{i=1}^{n} \left(x_i - x_i^*\right)^2 + \frac{\sum_{i=1}^{n} \left(y_i - a - bx_i^*\right)^2}{\lambda} \right) \to \min, \qquad (6)$$

где $\sigma^2 = \sigma_{\varepsilon_x}^2$.

Используя необходимые условия существования экстремума функции (6) $\frac{\partial S}{\partial x_i^*} = 0$, найдем

$$x_i^* = x_i + \frac{b}{\lambda + b^2}\left(y_i - a - bx_i\right), \; i = \overline{1,n}. \qquad (7)$$

Из условия $\frac{\partial S}{\partial a} = 0$, используя формулы (7), получим:

$$a = \overline{y} - b\overline{x}, \qquad (8)$$

где $\overline{y}$ и $\overline{x}$ — выборочные средние характеристики.

Из условия $\frac{\partial S}{\partial b} = 0$, используя формулы (7)–(8), получим квадратное уравнение [5, 14] относительно параметра b:

$$K_{xy} b^2 - \left(D_y - \lambda D_x\right) b - \lambda K_{xy} = 0, \qquad (9)$$

где D_x, D_y — выборочные дисперсии признаков; K_{xy} — ковариация. Смыслу задачи удовлетворяет только один из корней уравнения (9)

$$\hat{b} = \frac{\left(D_y - \lambda D_x\right) + \sqrt{\left(D_y - \lambda D_x\right)^2 + 4\lambda K_{xy}^2}}{2K_{xy}}. \qquad (10)$$

Таким образом для решения задачи (6) сначала найдем оценку параметра b по формуле (10), затем оценку параметра a по формуле (8) и, если требуется, определим оценки «истинных» значений объясняющей переменной по формулам (7). Очевидно, что значения этих оценок будут зависеть от заданного отношения дисперсий ошибок исследуемых признаков λ. При этом легко доказать, что регрессия Деминга является обобщением трех моделей регрессий [5, 8]:

- если $\lambda \to \infty$, то имеем прямую регрессию $y_i = a + bx_i$;
- если $\lambda = 1$, то регрессия является ортогональной;
- если $\lambda \to 0$, то получаем обратную регрессию $x_i = c + dy_i$.

Различные типы регрессий в зависимости от значений параметра λ представлены на рис. 1.

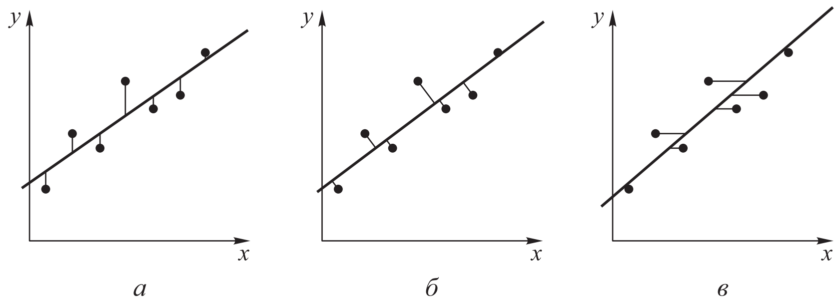

Рис. 1. Различные типы линий регрессии:
а — $\lambda \to \infty$; *б* — $\lambda = 1$; *в* — $\lambda \to 0$

Для прямой регрессии переменную y будем считать зависимой, а переменную x — независимой. При этом минимизируем сумму квадратов вертикальных расстояний от данных точек до линии регрессии (рис. 1, *а*). Для обратной регрессии, наоборот, x — зависимая, а y — независимая переменная. При этом минимизируем сумму квадратов горизонтальных расстояний от данных точек до линии регрессии (рис. 1, *в*). В ортогональной регрессии минимизируем сумму квадратов длин перпендикуляров, опущенных из данных точек на линию регрессии, т. е. минимизируем евклидово расстояние (рис. 1, *б*).

Помимо трех рассмотренных выделим еще один тип регрессии Деминга — диагональную [5]. Для диагональной регрессии отношение дисперсий ошибок переменных пропорционально отношению дисперсий наблюдаемых переменных, т. е. $\lambda = \dfrac{D_y}{D_x}$.

В работе [5] приведен аналитический вид смещений получаемых оценок параметров модели регрессии Деминга. Так, в предположении прямой регрессии, смещение оценки Деминга

$$\lim_{\lambda \to \infty} M(\hat{b}) = \frac{1}{1+\gamma_x^2} b,$$

где $\gamma_x^2 = \sigma_{\varepsilon_x}^2 / D_x$. При отсутствии ошибок по x положим, что $\gamma_x^2 = 0$, следовательно, оценка параметра b будет несмещенной. А если модель прямой регрессии оценить при наличии шума по x, то оценка параметра b будет занижена относительно его истинного значения.

В предположении обратной регрессии смещение оценки Деминга

$$\lim_{\lambda \to 0} M(\hat{b}) = \left(1 + \gamma_y^2\right) b,$$

где $\gamma_y^2 = \sigma_{\varepsilon_y}^2 / D_y$. Если отсутствуют ошибки по y, то оценка параметра b будет несмещенной. При наличии таких ошибок эта оценка будет приобретать смещение в сторону завышения.

В остальных случаях смещение оценки Деминга имеет вид

$$M(\hat{b}) = b\sqrt{\frac{1 + \gamma_y^2}{1 + \gamma_x^2}},$$

т. е. смещение зависит от уровней шума по y и x одновременно. В предположении диагональной регрессии $\gamma_x^2 = \gamma_y^2$, следовательно, оценка параметра b будет несмещенной.

Главная проблема при оценивании регрессии Деминга заключается в том, что заранее неизвестны дисперсии ошибок $\sigma_{\varepsilon_x}^2$ и $\sigma_{\varepsilon_y}^2$ исследуемых признаков, а значит, непонятно, какое именно значение параметра λ следует выбрать для моделирования.

Аналитический вид коэффициентов детерминации регрессии Деминга. Критериями оценки качества модели Деминга (4) будем считать коэффициент детерминации R_y^2 регрессии $y_i = a + bx_i^* + \varepsilon_{y_i}$ и коэффициент детерминации R_x^2 регрессии $x_i = x_i^* + \varepsilon_{x_i}$. В общем случае коэффициент детерминации R^2 показывает, какая доля дисперсии зависимой переменной трактуется дисперсией объясняющих переменных [3]:

$$R^2 = 1 - \frac{RSS}{TSS}, \tag{11}$$

где $RSS = \sum_{i=1}^{n} \varepsilon_i^2$ — остаточная сумма квадратов; $TSS = \sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2$ — полная сумма квадратов.

Коэффициент детерминации принимает значения в пределах $0 \le R^2 \le 1$. Чем выше R^2, тем выше качество регрессии.

Из уравнения $x_i = x_i^* + \varepsilon_{x_i}$ ошибка регрессии $\varepsilon_{x_i} = x_i - x_i^*$. Подставив в это выражение равенства (7)–(8), получим

$$\varepsilon_{x_i} = \frac{b^2\left(x_i - \overline{x}\right) - b\left(y_i - \overline{y}\right)}{\lambda + b^2}. \tag{12}$$

Возведем выражение (12) в квадрат и перейдем к сумме

$$\sum_{i=1}^{n} \varepsilon_{x_i}^2 = RSS_x = \frac{nb^2\left(D_x b^2 - 2K_{xy} b + D_y\right)}{\left(\lambda + b^2\right)^2}, \tag{13}$$

где D_x, D_y — выборочные дисперсии; K_{xy} — ковариация.

Представим равенство (13) в виде

$$RSS_x = \frac{nb\left(b^2\left(D_x b - K_{xy}\right) - \left(K_{xy} b^2 - D_y b\right)\right)}{\left(\lambda + b^2\right)^2}. \tag{14}$$

Из уравнения (9) имеем $K_{xy} b^2 - D_y b = \lambda\left(K_{xy} - D_x b\right)$. Тогда выражение (14) принимает вид

$$RSS_x = \frac{nb\left(b^2\left(D_x b - K_{xy}\right) + \lambda\left(D_x b - K_{xy}\right)\right)}{\left(\lambda + b^2\right)^2} = \frac{nb\left(D_x b - K_{xy}\right)}{\lambda + b^2}. \tag{15}$$

Учитывая, что по определению полная сумма квадратов $TSS_x = nD_x$, получим коэффициент детерминации

$$R_x^2 = 1 - \frac{RSS_x}{TSS_x} = \frac{nb\left(D_x b - K_{xy}\right)}{\left(\lambda + b^2\right) nD_x} = \frac{\lambda + b\left(K_{xy}/D_x\right)}{\lambda + b^2}. \tag{16}$$

Подставляя в выражение (16) соотношение (10), получим зависимость коэффициента детерминации R_x^2 регрессии $x_i = x_i^* + \varepsilon_{x_i}$ от параметра λ:

$$R_x^2 = \frac{K_{xy}^2\left(D_y + \lambda D_x + \sqrt{\left(D_y - \lambda D_x\right)^2 + 4\lambda K_{xy}^2}\right)}{D_x\left(\left(D_y - \lambda D_x\right)^2 + \left(D_y - \lambda D_x\right)\sqrt{\left(D_y - \lambda D_x\right)^2 + 4\lambda K_{xy}^2} + 4\lambda K_{xy}^2\right)}. \tag{17}$$

Для нахождения коэффициента детерминации R_y^2 предварительно подставим в выражение (7) равенство $x_i = x_i^* + \varepsilon_{x_i}$ и, учитывая, что $y_i - a - bx_i^* = \varepsilon_{y_i}$, получим соотношение $\lambda\varepsilon_{x_i} + b\varepsilon_{y_i} = 0$, из которого

$$\varepsilon_{y_i} = -\frac{\lambda}{b}\varepsilon_{x_i}. \qquad (18)$$

Возведем равенство (18) в квадрат и перейдем к сумме

$$\sum_{i=1}^{n}\varepsilon_{y_i}^2 = RSS_y = \frac{\lambda^2}{b^2}\sum_{i=1}^{n}\varepsilon_{x_i}^2. \qquad (19)$$

Подставив в выражение (19) соотношение (15), имеем

$$RSS_y = \frac{\lambda^2}{b^2}\frac{nb\left(D_x b - K_{xy}\right)}{\lambda + b^2} = \frac{n\lambda\left(D_y - K_{xy}b\right)}{\lambda + b^2}. \qquad (20)$$

Учитывая, что $TSS_y = nD_y$, получим коэффициент детерминации

$$R_y^2 = 1 - \frac{n\lambda\left(D_y - K_{xy}b\right)}{\left(\lambda + b^2\right)nD_y} = b\left(\frac{\lambda\left(K_{xy}/D_y\right) + b}{\lambda + b^2}\right). \qquad (21)$$

Подставив в выражение (21) соотношение (10), получим зависимость коэффициента детерминации R_y^2 регрессии $y_i = a + bx_i^* + \varepsilon_{y_i}$ от параметра λ:

$$R_y^2 = \left[\left(\lambda K_{xy}^2 + D_y^2 - \lambda D_x D_y\right)\left(D_y - \lambda D_x + \sqrt{\left(D_y - \lambda D_x\right)^2 + 4\lambda K_{xy}^2}\right) + 2\lambda K_{xy}^2 D_y\right] \Big/ \left[D_y\left(\left(D_y - \lambda D_x\right)^2 + \left(D_y - \lambda D_x\right)\sqrt{\left(D_y - \lambda D_x\right)^2 + 4\lambda K_{xy}^2} + 4\lambda K_{xy}^2\right)\right]. \qquad (22)$$

Задача максимизации качества регрессии Деминга. Предположим, что отсутствует информация о том, является ли модель регрессии Деминга прямой, обратной, ортогональной или диагональной. Зададимся целью выбрать такое значение параметра λ, чтобы качество регрессии Деминга по возможности было бы наибольшим. Для этого поставим следующую оптимизационную задачу:

$$L = R_y^2 + R_x^2 = 2 - \left(\frac{\sum_{i=1}^{n} \varepsilon_{x_i}^2}{D_x} + \frac{\sum_{i=1}^{n} \varepsilon_{y_i}^2}{D_y} \right) \to \max . \qquad (23)$$

Из сравнения функционалов S и L в задачах (5) и (23) следует, что при решении задачи (23) будут получены те же оценки параметров a и b, что и при решении задачи (5) при $\lambda = \frac{D_y}{D_x}$, т. е. в случае диагональной регрессии. Оценка параметра b диагональной регрессии [5] имеет вид

$$\hat{b} = \begin{cases} \sqrt{D_y / D_x}, \text{ если } K_{xy} > 0, \\ -\sqrt{D_y / D_x}, \text{ если } K_{xy} < 0. \end{cases} \qquad (24)$$

Если подставить в выражения для коэффициентов детерминации (17) и (22) значение $\lambda = \frac{D_y}{D_x}$, можно легко показать, что в точке экстремума функции (23) значения коэффициентов детерминации регрессий $x_i = x_i^* + \varepsilon_{x_i}$ и $y_i = a + bx_i^* + \varepsilon_{y_i}$ равны, т. е. $R_x^2 = R_y^2$.

Таким образом, для того чтобы качество регрессии Деминга было наилучшим с точки зрения максимизации функции (23), необходимо выбрать отношение дисперсий ошибок исследуемых признаков λ как отношение их выборочных дисперсий D_y / D_x. При этом оценки, полученные из решения задачи (23), будут несмещенными только в случае модели диагональной регрессии.

Пример численной обработки регрессии Деминга. Пусть в модели регрессии Деминга истинные значения параметров $a = 7$, $b = 12$, т. е. зависимость между переменными y^* и x^* имеет вид $y^* = 7 + 12x^*$. Для численной обработки такой регрессии были смоделированы 20 наблюдений, в которых переменная x^* принимает значения от 1 до 20. Для случайных остаточных составляющих ε_x и ε_y использовались случайные числа, распределенные по нормальному закону с нулевыми средними и постоянными дисперсиями: $\varepsilon_x \sim N\left(0, 1^2\right)$, $\varepsilon_y \sim N\left(0, 9^2\right)$. Используя соотношения (4), были вычислены значения переменных y и x. Результаты представлены в табл. 1.

Таблица 1

№ п/п	y^*	x^*	ε_y	ε_x	y	x
1	19	1	0,716	1,058	19,716	2,058
2	31	2	–3,847	–1,527	27,153	0,473
3	43	3	13,683	0,051	56,683	3,051
4	55	4	–0,244	0,918	54,756	4,918
5	67	5	–11,595	0,393	55,405	5,393
6	79	6	13,727	–0,789	92,727	5,211
7	91	7	–0,092	–1,398	90,908	5,602
8	103	8	10,819	–1,082	113,819	6,918
9	115	9	–8,986	0,119	106,014	9,119
10	127	10	1,682	–0,210	128,682	9,790
11	139	11	–1,492	–0,003	137,508	10,997
12	151	12	2,191	0,745	153,191	12,745
13	163	13	–7,273	–1,132	155,727	11,868
14	175	14	2,156	0,964	177,156	14,964
15	187	15	4,582	2,205	191,582	17,205
16	199	16	–3,839	–0,079	195,161	15,921
17	211	17	17,543	–0,054	228,543	16,946
18	223	18	5,633	0,580	228,633	18,580
19	235	19	9,042	0,719	244,042	19,719
20	247	20	1,913	–1,342	248,913	18,658

Найденные по исходным данным выборочные средние признаков $\overline{y} = 135,316$, $\overline{x} = 10,507$, выборочные дисперсии $D_y = 5031,773$, $D_x = 35,578$, выборочная ковариация $K_{xy} = 415,528$.

Графики зависимостей (17) и (22) коэффициентов детерминации от параметра λ представлены на рис. 2. Как видно, при $\lambda \to 0$ имеем обратную регрессию, а при $\lambda \to \infty$ — прямую регрессию. Решением задачи (23) является точка пересечения графиков функций, имеющая координаты $\lambda = D_y/D_x = 5031,773 / 35,578 = 141,43$ и $R^2 = 0,991$. По данным графиков видим, что если взять значение параметра λ больше, чем 141,43, то уменьшатся ошибки по переменной x, но увеличатся ошибки по переменной y, а если взять λ меньше, чем 141,43, — то наоборот.

Результаты оценивания параметров модели регрессии Деминга в предположении о ее виде представлены в табл. 2.

Согласно данным таблицы, все оценки параметра $a = 7$ оказались завышенными. Оценки параметра $b = 12$ в случае обратной и ортогональной регрессии оказались завышенными, а в случае прямой и диа-

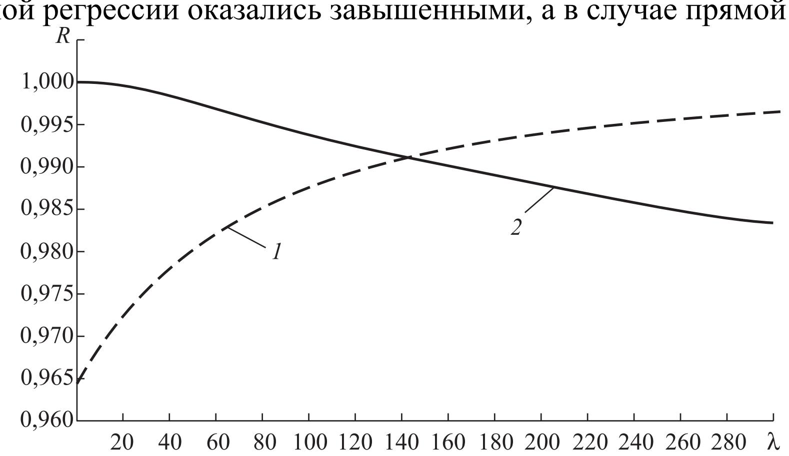

Рис. 2. Зависимости коэффициентов детерминации от параметра λ:
1 — коэффициент детерминации R_x; *2* — коэффициент детерминации R_y

Таблица 2

Предположение	λ	a	b	R_x	R_y	R_x+R_y
Прямая	1 000 000	12,603	11,679	1	0,964	1,964
Обратная	0	8,085	12,109	0,964	1	1,964
Ортогональная	1	8,117	12,106	0,965	1	1,965
Диагональная	141,43	10,365	11,892	0,991	0,991	1,982

гональной регрессии — заниженными. При этом максимальное значение суммы коэффициентов детерминации дает диагональная регрессия.

Очевидно, что одного эксперимента такого типа недостаточно для оценки качества параметров регрессии. Поэтому с помощью метода Монте-Карло [15] была проведена серия из ста тысяч экспериментов. В каждой реализации была получена выборка, и по ней вычислены оценки коэффициентов a и b. Затем по всем ста тысячам реализациям получены средние значения оценок коэффициентов. Результаты показаны в табл. 3.

Таблица 3

Предположение	Среднее значение a	Среднее значение b
Прямая	10,135	11,702
Обратная	4,86	12,204
Ортогональная	4,927	12,198
Диагональная	7,558	11,947

Таким образом, все оценки параметров оказались смещенными. Эти результаты полностью согласуются с результатами, полученными в работе [5]. Так, при оценивании прямой регрессии при наличии шума по переменной x оценка параметра b занижается относительно его истинного значения. При оценивании обратной регрессии оценка приобретает смещение в сторону завышения. Оценки диагональной регрессии в нашем примере оказались наименее смещенными.

Выводы. Получены аналитические зависимости коэффициентов детерминации от соотношения дисперсий ошибок в модели регрессии Деминга. Определено такое соотношение дисперсий ошибок исследуемых признаков в модели регрессии Деминга, которое дает максимальную величину суммы коэффициентов детерминации. Рассмотрен модельный пример численной обработки регрессии Деминга в предположениях о том, что регрессия является прямой, обратной, ортогональной и диагональной. Доказана смещенность полученных в результате оценивания параметров.

ЛИТЕРАТУРА

[1] Айвазян С.А. *Прикладная статистика. Т. 2: Основы эконометрики*. Москва, ЮНИТИ-ДАНА, 2001, 432 с.
[2] Носков С.И. *Технология объектов с нестабильным функционированием и неопределенностью в данных*. Иркутск, Облинформпечать, 1996, 321 с.
[3] Гефан Г.Д. *Эконометрика*. Иркутск, Изд-во ИрГУПС, 2005, 84 с.
[4] Дрейпер Н., Смит Г. *Прикладной регрессионный анализ*. Москва, Финансы и статистика, 1986, 366 с.
[5] Тимофеев В.С., Щеколдин В.Ю., Тимофеева А.Ю. Идентификация зависимостей признаков стохастической природы на основе регрессии Деминга. *Информатика и ее применения*, 2013, т. 7, № 2, с. 60–68.
[6] Deming W.E. *Statistical adjustment of data*. New York, Dover, 2011, 288 p.
[7] Besalu E., de Julian-Ortiz J., Pogliani L. Ordinary and orthogonal regressions in QSAR/QSPR and chemistry-related studies. *Match-Communications in Mathematical and in Computer Chemistry*, 2010, no. 63, pp. 573–583.
[8] Huang L., Rattner A., Liu H., Nathans J. Tutorial: how to draw the line in biomedical research. *eLife*, 2013. Available at: https://elifesciences.org/content/2/e00638
[9] Bernard G., Bernadette B. Measurement methods comparison with errors-in-variables regressions. From horizontal to vertical OLS regression, review and new perspectives. *Chemometrics and Intelligent Laboratory Systems*, 2014, vol. 134, pp. 123–139. DOI 10.1016/j.chemolab.2014.03.006
[10] Dhanoa M.S., Sanderson R., Lopez S., Dijkstra J., Kebreab E., France J. Regression procedures for relationships between random variables. *Modelling Nutrient Digestion and Utilisation in Farm Animals*, 2011, part 1, pp. 31–39. DOI 10.3920/978-90-8686-712-7_3
[11] Shaoji Xu. A Property of Geometric Mean Regression. *The American Statistician*, 2014, vol. 68, iss. 4, pp. 277–281. DOI 10.1080/00031305.2014.962763

[12] Haeckel R., Wosniok W., Klauke R. Comparison of ordinary linear regression, orthogonal regression, standardized principal component analysis, Deming and Passing-Bablok approach for method validation in laboratory medicine. *Laboratoriums Medizin*, 2013, vol. 37, iss. 3, pp. 147–163. DOI 10.1515/labmed-2013-0003
[13] Kallner A. Comprehensive method comparisons: getting more from the data. *Accreditation and Quality Assurance*, 2014, vol. 19, iss. 6, pp. 451–457. DOI 10.1007/s00769-014-1089-9
[14] Glaister P. Least squares revisited. *The Mathematical Gazette*, 2001, no. 85, pp. 104–107.
[15] Доугерти К. *Введение в эконометрику*. Москва, ИНФРА-М, 2009, 465 с.

Статья поступила в редакцию 18.05.2016

Ссылку на эту статью просим оформлять следующим образом:
Базилевский М.П. Аналитические зависимости между коэффициентами детерминации и соотношением дисперсий ошибок исследуемых признаков в модели регрессии Деминга. *Математическое моделирование и численные методы*, 2016, № 2 (10), с. 104–116.

Базилевский Михаил Павлович — канд. техн. наук, доцент кафедры «Математика» ИрГУПС. e-mail: mik2178@yandex.ru

Analytical dependences between the determination coefficients and the ratio of error variances of the test items in Deming regression model

© M.P. Bazilevskiy

Irkutsk State Transport University, Irkutsk, 664074, Russia

The article deals with the problem of building regression models, in which all variables have stochastic nature. To solve it, we propose to use the determination coefficient. We obtain analytical dependencies of the determination coefficients from the ratio of error variances of the test items. We set the optimization problem, assuming the maximization of the determination coefficients sum for each Deming regression equation. We give a model example of the numerical processing of Deming regression with its parameters and sign errors which are known.

Keywords: *Deming regression model, least square method, error variance, determination coefficient.*

REFERENCES

[1] Aivazyan S.A. *Prikladnaia statistika. V. 2: Osnovy ekonometriki* [Applied statistics. Vol. 2: Fundamentals of econometrics]. Moscow, YuNITI-DANA Publ., 2001, 432 p.
[2] Noskov S.I. *Tekhnologiya obyektov s nestabilnym funktsionirovaniem i neopredelennostyu v dannykh* [Technology of the objects with unstable operation and the data uncertainty]. Irkutsk, Oblinformpechat Publ., 1996, 321 p.
[3] Gefan G.D. *Ekonometrika* [Econometrics]. Irkutsk, IrGUPS Publ., 2005, 84 p.

[4] Dreiper N., Smit G. *Prikladnoi regressionnyi analiz* [Applied regression analysis]. Moscow, Finansy i statistika Publ., 1986, 366 p.
[5] Timofeev V.S., Shchekoldin V.Yu., Timofeeva A.Yu. *Informatika i ee primeneniya — Informatics and Applications,* 2013, vol. 7, no. 2, pp. 60–68.
[6] Deming W.E. *Statistical adjustment of data*. New York, Dover, 2011, 288 p.
[7] Besalu E., de Julian-Ortiz J., Pogliani L. *Match-Communications in Mathematical and in Computer Chemistry*, 2010, no. 63, pp. 573–583.
[8] Huang L., Rattner A., Liu H., Nathans J. Tutorial: how to draw the line in biomedical research. *eLife*, 2013. Available at: https://elifesciences.org/content/2/e00638
[9] Bernard G., Bernadette B. *Chemometrics and Intelligent Laboratory Systems*, 2014, vol. 134, pp. 123–139. DOI 10.1016/j.chemolab.2014.03.006
[10] Dhanoa M.S., Sanderson R., Lopez S., Dijkstra J., Kebreab E., France J. *Modelling Nutrient Digestion and Utilisation in Farm Animals*, 2011, part 1, pp. 31–39. DOI 10.3920/978-90-8686-712-7_3
[11] Shaoji Xu. *The American Statistician*, 2014, vol. 68, iss. 4, pp. 277–281. DOI 10.1080/00031305.2014.962763
[12] Haeckel R., Wosniok W., Klauke R. *Laboratoriums Medizin*, 2013, vol. 37, iss. 3, pp. 147–163. DOI 10.1515/labmed-2013-0003
[13] Kallner A. *Accreditation and Quality Assurance*, 2014, vol. 19, iss. 6, pp. 451–457. DOI 10.1007/s00769-014-1089-9
[14] Glaister P. *The Mathematical Gazette*, 2001, no. 85, pp. 104–107.
[15] Dougerti K. *Vvedenie v ekonometriku* [Introduction to econometrics]. Moscow, INFRA-M Publ., 2009, 465 p.

Bazilevskiy M.P., Cand. Sci. (Eng.), Assoc. Professor of the Mathematics Department at the Irkutsk State Transport University. e-mail: mik2178@yandex.ru