УДК 519.281: 528.11: 004.9

Джунь Йосип Володимирович д.ф.-м.н., професор, завідувач кафедри математичного моделювання, josif-june@rambler.ru, **Лотюк Юрій Георгійович,** к.пед.н., доцент, доцент кафедри математичного моделювання (Міжнародний економіко-гуманітарний університет імені академіка Степана Дем'янчука), lotyuk@ukr.net

АЛГОРИТМ ВАЖЛИВОЇ ПРОЦЕДУРИ ПЕРЕД ЗАПУСКОМ ПРОГРАМИ ОБЧИСЛЕНЬ ПАРАМЕТРІВ РЕГРЕСІЙНОЇ МОДЕЛІ

Кожен метод математичного моделювання ґрунтується на тих чи інших фундаментальних положеннях, виконання яких є необхідним для його коректного застосування. Параметри регресійної моделі визначається методом квадратів (МНК), якій розроблений математичними геніями А. М. Лежандром (1806 р.), К. Ф. Гаусом (1809 р.) і П. С. Лапласом (1812 р.) [1–3]. Програмісти, які використовують ті чи інші програмні продукти з регресійного аналізу (РА), як правило, вважають, що ці видатні математики добре розібралися у запропонованому ними МНК і ніяких сумнівів в тому, чи можна його застосовувати до наявних даних, бути не може. Такий підхід свідчить про невігластво виконавця. По-перше тому, що вихідні дані для регресійного аналізу можуть бути неякісними, зробленими халтурно або просто непрофесійно. А звинувачення потім падають на програміста, якій, мовляв, не перевірив неякісні дані [4]. По-друге, навіть якісно отримані вихідні дані можуть не відповідати фундаментальним положенням (РА) [5]. Тому перед застосуванням програмного продукту з РА необхідно спочатку перевірити, чи відповідають наявні дані необхідним умовам застосування регресійної моделі, загальний вигляд якої є наступним:

$$y_i = a_0 + a_1 x_{1i} + a_2 x_{2i} + \cdots + a_k x_{ki} + e_i \ , \qquad (1)$$

де y_i – значення досліджуваної ознаки; x_{ij} – значення i-го факторного впливу на спостереження $y_i (j = 1,2, \dots, k; i = 1,2, \dots, n)$; $a_0, a_1, \dots, a_k$ – регресори, обчислення яких і є метою РА; e_i – залишкові похибки; також потрібно пам'ятати, щоб $n \geq 30$ для того щоб отримати параметри моделі які мали б належне статистичне обґрунтування [5].

Перше ніж почати РА необхідно, принаймні перевірити головну умову його застосування, а саме те, що залежність кожної з пар:

$$y_i \rightarrow x_{ji}, \quad (2)$$

мусить мати лінійний характер, тобто необхідно перевірити, що всі кореляційні поля залежностей (2) мають цілком очевидну еліпсоїдальність.

Нехай для прикладу, ми маємо пари даних $y_i \rightarrow x_{1i}$, тобто, досліджуємо лінійність залежності y_i і значень x_{1i} для першого фактора моделі (1) (табл. 1).

Таблиця 1

Вихідні дані для побудови кореляційного поля залежності $y_i \rightarrow x_{1i}$.

i	y_i	x_{1i}	i	y_i	x_{1i}	i	y_i	x_{1i}	i	y_i	x_{1i}
1	6.00	5.00	9	9.55	22.15	17	20.00	45.12	25	18.21	65.05
2	7.50	7.47	10	16.00	27.50	18	11.00	47.47	26	21.71	67.65
3	5.90	9.90	11	9.00	29.89	19	16.09	50.51	27	18.71	69.69
4	9.10	12.60	12	13.83	32.40	20	14.05	52.52	28	19.35	72.35
5	6.14	15.15	13	10.99	35.05	21	16.06	54.87	29	21.15	74.93
6	11.60	17.55	14	19.00	37.38	22	21.00	57.59	30	24.50	77.76
7	8.08	19.99	15	14.07	39.92	23	15.00	60.00	31	22.92	80.01
8	11.33	22.39	16	10.90	42.50	24	20.88	62.60	32	25.01	82.57

Нам необхідно перевірити лінійність зв'язку у парах $y_i \rightarrow x_{1i}$. Алгоритм вирішення цієї задачі в середовищі MathCad наступний:

1. Вносимо в комп'ютер всю базу вихідних даних у вигляді (2), які необхідні для регресійного аналізу.

2. Вводимо в комп'ютер програму побудови кореляційного поля залежностей (2) в прямокутній системі координат.

3. Будуємо кореляційне поле залежності $y_i = x_{1i}$, тобто, наносимо прямокутній системі координат усі точки з координатами $\{x_{1i}, y_i\}$ на графік і зберігаємо його.

Далі будуємо таким же чином на новому графіку кореляційне поле залежності $y_i = x_{2i}$ і зберігаємо його. Аналогічним чином будуємо і зберігаємо кореляційні поля зроблені для всіх пар (2).

4. Роздруковуємо усі кореляційні поля уважно їх аналізуємо.

Висновки: Якщо усі отримані кореляційні поля залежностей $y_i = x_{ji}$ для дії усіх факторів j мають цілком очевидну еліпсоїдність, як показано на рис. 1 або принаймні групуються навколо якоїсь прямої, то це є свідченням лінійності моделі і можна сміливо починати регресійний аналіз. Але якщо хоча б у деякої однієї пари у формулі (2) кореляційне поле не є еліпсоїдальним і по своїй формі явно далеке від нього, то регресійна модель має порушену лінійність і створювати модель не можна.

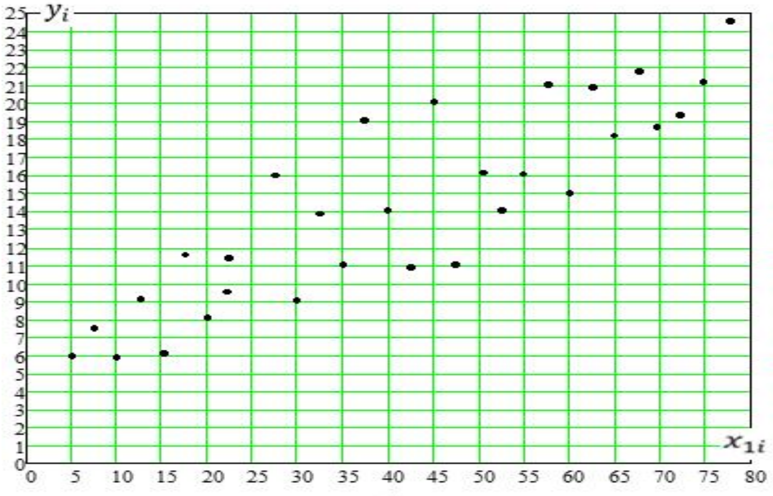

Рис. 1. Кореляційне поле залежності $y_i \to x_{1i}$

В цьому разі необхідно привести спочатку цю залежність до лінійного вигляду шляхом заміни змінних, або застосувати комп'ютерні програми лінеаризації нелінійної залежності. Після цього можна починати РА і обчислювати регресори.

Побудова кореляційних полів є ефективним засобом попередньої перевірки лінійності регресійної моделі тим більше коли вона виконується на комп'ютері по рекомендованому нам алгоритму. Ця процедура дозволяє досить простими засобами забезпечити коректність моделювання.

Список використаних літературних джерел

1. Legendre A. M. Nouvelles methods pour la determination des orbites des comets. Appendice sur la method des moindres carres. Paris. 1806.

2. Gauss C. F. Theoria motus corporum coelestium in sectionibus conicis Solet ambientium. Hamburg, 1809.

3. Laplace P. S. Theory analytique des probabilities. Paris: Coursier, 1812.

4. Атраментова Л. А. Наукове дослідження і статистика. *Науковий світ,* 2006. № 4. С. 6–7.

5. Джунь Й. В. Неклассическая теория погрешностей измерений. Ровно: Естеро, 2015. 168 с.