Журнал обчислювальної та прикладної математики | 2016, №3(123) | Journal of Computational & Applied Mathematics

УДК 517.5
MSC 517.5

CHECKING OF POLINOMIAL REGRESSION MODEL FOR CORRECTNESS

Marta Savkina
Institute of Mathematics of NASU, Kyiv, Ukraine, E-mail: marta@kiev.ua

ПЕРЕВІРКА ПОЛІНОМІАЛЬНОЇ МОДЕЛІ РЕГРЕСІЇ НА КОРЕКТНІСТЬ

М. Ю. Савкіна
Інститут математики НАНУ, Київ, Україна, E-mail: marta@kiev.ua

Abstract. The problem of exclusion of variables in polynomial regression model is examined that arises in the selection of model. In the case of the uniform arrangement of observation points sign of equality to zero OLS estimation of parameters of quadratic model as a equality to zero of linear combination of values of regressand with constant coefficients is obtained. This sign can be used for testing hypothesis of equality to zero of unknown regression parameter. Testing of this hypothesis is carried out by using the likelihood ratio criterion, which determines to accept the hypothesis if corresponding linear combination less some positive number. Otherwise a criterion rejects a hypothesis. This number for all three parameters of quadratic model is found; it depends on the level of meaningfulness, with that a hypothesis will be accepted, from residual sum of squares of rejections and from dispersion of estimation examined parameter.

Such sign is obtained for the OLS estimation of leading coefficient of polynomial regression also at the paper. The coefficient at the value of regressand y_i in linear combination is presented as a polynomial of degree k from discrete variable i, where k - degree of leading coefficient of polynomial regression. The coefficients of this polynomial depend from k and from n, where $n + 1$ - is number of observation points. They can be presented as a decision of the system of linear algebraic equalizations of the special kind. Elements of matrix of this system are sums of squares of degrees of natural numbers, they depend from n. An algorithm that allows to express the decision of the system of order k through the decision of two systems of order $k - 1$ with identical matrices and different columns of free terms is built. The matrix of the systems of order $k - 1$ is analogical to the matrix of the system of order to k. Due to this algorithm it is possible to present decision of the system of order k as formulas at arbitrary n. For this purpose it is needed to write the decisions k systems of order 1 (it is simple fractions, the numerator of that is sum of squares of degrees of natural numbers, the denominator is number of observation points),further, it is needed to write the decisions $k-1$ systems of order 2 and so on, two systems of order $k-1$ and in conclusion decision of the desired system.

For $k = 3$ and $k = 4$ the coefficients at the value of regressand y_i, $i = 0, 1, ..., n$, are found in an obvious kind; they are presented as rational fractions from to the variable of n.
Keywords: OLS estimation, regressand, polynomial regression.

Резюме. Розглядається задача виключення змінних в поліноміальній моделі регресії, що виникає при підборі моделі. У випадку рівномірного розташування точок спостереження отримано ознаку рівності нулю оцінки МНК параметрів квадратичної моделі у вигляді рівності нулю лінійної комбінації значень відклику зі сталими коефіцієнтами. Цю ознаку можна використовувати для перевірки гіпотези рівності нулю невідомого параметра регресії. У роботі також отримано таку ознаку для оцінки МНК старшого коефіцієнту поліноміальної регресії, побудовано алгоритм для знаходження коефіцієнтів відповідної лінійної комбінації.
Ключові слова: оцінка МНК, відклик, поліноміальна регресія.

Вступ

Класичний регресійний аналіз засновано на тому, що вигляд моделі регресії відомий з точністю до параметрів, тобто, набір незалежних змінних (факторів) задано однозначно, всі істотні змінні присутні та ніяких альтернативних способів вибору факторів немає. Насправді тісно пов'язаний з вибором моделі об'єкта вибір регресорів — одна з найскладніших проблем. В середині минулого сторіччя поява ЕОМ значно спростила цю проблему. „...Поступово з'ясувалося, що ЕОМ допускає відмову від жорсткої моделі дослідження та підбір під час обробки даних деякої „найкращої“ моделі...“ [1]. У даний час розроблено багато статистичних методів відбору змінних, такі як метод всіх можливих регресій, метод виключення, кроковий регресійний метод, ступінчастий регресійний аналіз, гребенева регресія тощо. В одній і тій ж задачі їх застосування не завжди веде до отримання тієї ж самої моделі, хоча в багатьох випадках виходить однаковий результат. Кожен метод має свої недоліки, свої переваги. Жоден метод не буде добре працювати в усіх випадках. Використання цих методів для розв'язання практичних задач можна знайти наприклад в роботах [5], [6], [7].

У деяких з цих методів застосовується критерій Фішера для перевірки гіпотези рівності нулю невідомого параметра регресії, і на його підставі приймається рішення про видалення відповідного фактора з регресії.

1. Умови рівності нулю параметрів моделі квадратичної регресії

Розглянемо модель квадратичної регресії

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i, \qquad i = 0, 1, ..., n, \tag{1}$$

де $\epsilon_0, ..., \epsilon_n$ — незалежні у сукупності нормально розподілені випадкові величини з $E\epsilon_i = 0$ та $D\epsilon_i = \sigma^2$, $\beta_0, \beta_1, \beta_2$ — невідомі параметри моделі, які підлягають оцінюванню.

Найпоширенішим методом оцінювання невідомих параметрів у регресійному аналізі є метод найменших квадратів (МНК), який полягає в мінімізації суми

$$\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1 x_i - \beta_2 x_i^2)^2$$

відносно $\beta_0, \beta_1, \beta_2$.

Для знаходження оцінок МНК $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ параметрів $\beta_0, \beta_1, \beta_2$ маємо наступну систему лінійних алгебраїчних рівнянь

$$\begin{cases} \sum_{i=0}^{n}\big(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i - \hat{\beta}_2 x_i^2\big) = 0, \\ \sum_{i=0}^{n}\big(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i - \hat{\beta}_2 x_i^2\big)x_i = 0, \\ \sum_{i=0}^{n}\big(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i - \hat{\beta}_2 x_i^2\big)x_i^2 = 0, \end{cases} \tag{2}$$

з якої отримуємо

$$\hat{\beta}_2 = \frac{1}{\Delta}\sum_{i=0}^{n}\Big\{\sum_{j=0}^{n}(x_j^2 - \bar{x}^2)\Big(x_i^2 - \frac{\sum_{k=0}^{n} x_k^2}{n}\Big) - \sum_{j=0}^{n} x_j^2(x_j - \bar{x})(x_i - \bar{x})\Big\}y_i,$$

$$\Delta = \sum_{j=0}^{n}(x_j^2 - \bar{x}^2)\sum_{j=0}^{n} x_j^2\Big(x_j^2 - \frac{\sum_{k=0}^{n} x_k^2}{n}\Big) - \Big[\sum_{j=0}^{n} x_j^2(x_j - \bar{x})\Big]^2, \tag{3}$$

$$\bar{x} = \frac{1}{n+1}\sum_{j=0}^{n} x_j;$$

$$\hat{\beta}_1 = \frac{1}{\sum_{j=0}^{n}(x_j^2 - \bar{x}^2)}\Big(\sum_{i=0}^{n} y_i(x_i - \bar{x}) - \hat{\beta}_2\sum_{j=0}^{n} x_j^2(x_j - \bar{x})\Big); \tag{4}$$

$$\hat{\beta}_0 = \bar{y} - \bar{x}\hat{\beta}_1 - \frac{\hat{\beta}_2\sum_{k=0}^{n} x_k^2}{n+1}; \tag{5}$$

$$\bar{y} = \frac{1}{n+1}\sum_{j=0}^{n} y_j.$$

Поставимо задачу визначити, дорівнює нулю оцінка МНК якогось параметра квадратичної моделі (наприклад, $\hat{\beta}_2$) чи ні. Звичайно, можна розв'язати систему (2) та знайти $\hat{\beta}_2$, одночасно з $\hat{\beta}_1$ та $\hat{\beta}_0$. Але в деяких випадках можна знайти умову рівності нулю оцінки МНК параметра $\hat{\beta}_2$, перевірити яку буде значно легше, ніж розв'язати систему рівнянь.

Теорема 1. *Нехай $x_i = \frac{i}{n}, i = 0, 1, ..., n$. Тоді*

$$\hat{\beta}_2 = 0 \;\Leftrightarrow\; L_2(y_0, y_1, ..., y_n) = \sum_{i=0}^{n} a_i^{(2)} y_i = 0, \tag{6}$$

$$\hat\beta_1 = 0 \;\Leftrightarrow\; L_1(y_0, y_1, ..., y_n) = \sum_{i=0}^{n} a_i^{(1)} y_i = 0, \tag{7}$$

$$\hat\beta_0 = 0 \;\Leftrightarrow\; L_0(y_0, y_1, ..., y_n) = \sum_{i=0}^{n} a_i^{(0)} y_i = 0, \tag{8}$$

де

$$a_i^{(2)} = \Big(i - \frac{n}{2}\Big)^2 - \frac{n(n+2)}{12},$$

$$a_i^{(1)} = \Big(i - \frac{(2n+1)(8n-3)}{30n}\Big)^2 - \frac{(2n+1)(-38n^3 - 58n^2 + 30n - 9)}{900n^4},$$

$$a_i^{(0)} = \Big(i - \frac{3(2n+1)}{10n}\Big)^2 - \frac{6n^2+6n-11}{100n^2}.$$

Доведення. У випадку $x_i = \frac{i}{n}, i = 0, 1, ..., n$ формули (3), (4), (5) після низки перетворень приймуть вигляд

$$\hat\beta_2 = \frac{1}{\Delta}\sum_{i=0}^{n} \frac{(n+1)(n+2)}{12n}\Big(x_i^2 - x_i + \frac{n-1}{6n}\Big)y_i,$$

$$\hat\beta_1 = \frac{1}{\Delta}\sum_{i=0}^{n} \frac{(n+1)(n+2)}{12n}\Big(x_i^2 - \frac{(2n+1)(8n-3)}{15n^2}x_i + \frac{(2n+1)(n-1)}{10n^2}\Big)y_i,$$

$$\hat\beta_0 = \frac{1}{\Delta}\sum_{i=0}^{n} \frac{(n+1)(n+2)(n-1)}{72n^2}\Big(x_i^2 - \frac{6(2n+1)}{10n}x_i + \frac{3n^2+3n+2}{10n^2}\Big)y_i.$$

Ураховуючи, що в даному випадку

$$\Delta = \frac{(n+1)^2(n+2)^2(n-1)(n+3)}{2160n^4},$$

маємо

$$\hat\beta_2 = \frac{180n^3}{(n+1)(n+2)(n-1)(n+3)}\sum_{i=0}^{n}\bigg(\Big(i - \frac{n}{2}\Big)^2 - \frac{n(n+2)}{12}\bigg)y_i, \tag{3'}$$

$$\hat\beta_1 = \frac{180n^3}{(n+1)(n+2)(n-1)(n+3)}\times$$
$$\times\sum_{i=0}^{n}\bigg(\Big(i - \frac{(2n+1)(8n-3)}{30n}\Big)^2 - \frac{(2n+1)(-38n^3-58n^2+30n-9)}{900n^4}\bigg)y_i, \tag{4'}$$

$$\hat\beta_0 = \frac{30n^2}{(n+1)(n+2)(n+3)}\sum_{i=0}^{n}\bigg(\Big(i - \frac{3(2n+1)}{10n}\Big)^2 - \frac{6n^2+6n-11}{100n^2}\bigg)y_i. \tag{5'}$$

З формул $(3'), (4'), (5')$ випливають твердження (6), (7), (8).
Теорему доведено.
З теореми 1 для кокретних n отримуємо такі співвідношення
$n = 3$

$$\hat{\beta}_2 = 0 \;\Leftrightarrow\; y_0 - y_1 - y_2 + y_3 = 0;$$
$$\hat{\beta}_1 = 0 \;\Leftrightarrow\; 21y_0 - 13y_1 - 17y_2 + 9y_3 = 0;$$
$$\hat{\beta}_0 = 0 \;\Leftrightarrow\; 19y_0 + 3y_1 - 3y_2 + y_3 = 0,$$

$n = 4$

$$\hat{\beta}_2 = 0 \;\Leftrightarrow\; 2y_0 - y_1 - 2y_2 - y_3 + 2y_4 = 0;$$
$$\hat{\beta}_1 = 0 \;\Leftrightarrow\; 54y_0 - 13y_1 - 40y_2 - 27y_3 + 26y_4 = 0;$$
$$\hat{\beta}_0 = 0 \;\Leftrightarrow\; 31y_0 + 9y_1 - 3y_2 - 5y_3 + 3y_4 = 0,$$

$n = 5$

$$\hat{\beta}_2 = 0 \;\Leftrightarrow\; 5y_0 - y_1 - 4y_2 - 4y_3 - y_4 + 5y_5 = 0;$$
$$\hat{\beta}_1 = 0 \;\Leftrightarrow\; 165y_0 - y_1 - 92y_2 - 108y_3 - 49y_4 + 85y_5 = 0;$$
$$\hat{\beta}_0 = 0 \;\Leftrightarrow\; 46y_0 + 18y_1 - 8y_3 - 6y_4 + y_5 = 0$$

і так далі.

2. Задача перевірки моделі на коректність

Теорема 1 дає умову рівності нулю оцінки МНК параметрів квадратичної моделі у вигляді рівності нулю лінійної комбінації значень відклику зі сталими коефіцієнтами при конкретному n. Але на практиці така лінійна комбінація майже ніколи точно дорівнювати нулю не буде. Проте цю теорему можна використовувати в задачах побудови моделі регресії. Складовою частиною таких задач є перевірка гіпотези рівності нулю невідомого параметра регресії. У випадках, коли така гіпотеза підтверджується з великою ймовірністю, фактор як правило виключають з регресії, якщо ні — залишають.

У зв'язку з цим розглянемо задачу перевірки простої статистичної гіпотези

$$H_2 : \; \beta_2 = 0.$$

Критерій відношення правдоподібності перевірки гіпотези приводить до множини прийняття гіпотези H_2 [2]

$$E_{H_2} = \{y \in R^{n+1} : \frac{S_{2,2}^2 - S_3^2}{S_3^2} \leq \varphi\},$$

де

$$S_3^2 = \sum_{i=0}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i - \hat{\beta}_2 x_i^2)^2,$$

$$S_{2,2}^2 = \sum_{i=0}^{n} (y_i - \hat{\beta}_{0,2} - \hat{\beta}_{1,2} x_i)^2,$$

$\hat{\beta}_{0,2}, \hat{\beta}_{1,2}$ — оцінки МНК невідомих параметрів моделі лінійної регресії

$$y_i = \beta_{0,2} + \beta_{1,2} x_i + \epsilon_i, \qquad i = 0, 1, ..., n.$$

Значення φ в лінійній регресії вибирається так [2]. Припустимо, модель лінійній регресії має m невідомих параметрів, а гіпотеза H складається з p лінійних рівнянь. Задамося рівнем значущості α та знайдемо для нього відповідне значення $F_\alpha(p, n+1-m)$. Позначимо через $f(t,p,n+1-m)$ щільність розподілу Фішера з p та $n+1-m$ ступенями свободи та знайдемо таке $F_\alpha = F_\alpha(p, n+1-m)$, що $\int_{F_\alpha}^{\infty} f(t,p,n+1-m)dt = \alpha$. Значення F_α знаходять з таблиць. Покладемо

$$\varphi = \frac{p}{n+1-m} F_\alpha(p, n+1-m).$$

У нашому випадку $m = 3$ (модель (1) має 3 невідомі параметри) та $p = 1$ (гіпотеза H_2 складається з одного рівняння), тобто

$$\varphi = \frac{1}{n-2} F_\alpha(1, n-2).$$

Позначимо через z_i, $i = 0,1,2$ — i-й діагональний елемент матриці $(X'X)^{-1}$.

Згідно з [3]

$$S_{2,2}^2 = \sum_{i=0}^{n} (y_i - \bar{y})^2 - \frac{(\sum_{i=0}^{n} y_i(x_i - \bar{x}))^2}{\sum_{i=0}^{n} (x_i - \bar{x})^2},$$

а різницю $S_{2,2}^2 - S_3^2$ можна подати у вигляді [2]

$$S_{2,2}^2 - S_3^2 = \frac{\hat{\beta}_2^2}{z_2}. \tag{9}$$

Далі, у випадку $x_i = \frac{i}{n}, i = 0,1,...,n$

$$z_2 = \frac{180n^3}{(n+1)(n+2)(n-1)(n+3)},$$

а з $(3')$ маємо

$$\hat{\beta}_2 = z_2 L_2(y_0, y_1, ..., y_n).$$

Отже, множина прийняття гіпотези H_2 набирає вигляду

$$E_{H_2} = \left\{ y \in R^{n+1} : L_2^2(y_0, y_1, ..., y_n) \le \frac{S_3^2 \varphi}{z_2} \right\}.$$

S_3^2 знаходимо з формули (9).

Розглянемо тепер задачу перевірки гіпотези

$$H_1: \ \beta_1 = 0.$$

Критерій відношення правдоподібності приводить до множини прийняття гіпотези H_1

$$E_{H_1} = \{ y \in R^{n+1} : \frac{S_{2,1}^2 - S_3^2}{S_3^2} \le \varphi \},$$

де

$$S_{2,1}^2 = \sum_{i=0}^{n} (y_i - \hat\beta_{0,1} - \hat\beta_{2,1} x_i^2)^2,$$

$\hat\beta_{0,1}, \hat\beta_{2,1}$– оцінки МНК невідомих параметрів моделі лінійної регресії

$$y_i = \beta_{0,1} + \beta_{2,1} x_i^2 + \epsilon_i, \qquad i = 0, 1, ..., n.$$

Після низки перетворень маємо

$$S_{2,1}^2 = \sum_{i=0}^{n} (y_i - \bar y)^2 - \frac{\left(\sum_{i=0}^{n} y_i \left(\frac{\sum_{j=0}^{n} x_j^2}{n+1} - x_i^2\right)\right)^2}{\sum_{i=0}^{n} \left(\frac{\sum_{j=0}^{n} x_j^2}{n+1} - x_i^2\right)^2},$$

а різницю $S_{2,1}^2 - S_3^2$ можна подати у вигляді [2]

$$S_{2,1}^2 - S_3^2 = \frac{\hat\beta_1^2}{z_1}.$$

У випадку $x_i = \frac{i}{n}, i = 0, 1, ..., n$

$$z_1 = \frac{12n(2n+1)(8n-3)}{(n+1)(n+2)(n-1)(n+3)},$$

а з $(4')$ маємо

$$\hat\beta_1 = z_1 L_1(y_0, y_1, ..., y_n).$$

Множина прийняття гіпотези H_1 набирає вигляду

$$E_{H_1} = \left\{ y \in R^{n+1} : L_1^2(y_0, y_1, ..., y_n) \le \frac{S_3^2 \varphi}{z_1} \right\}.$$

Насамкінець розглянемо задачу перевірки гіпотези

$$H_0 : \ \beta_0 = 0.$$

Критерій відношення правдоподібності приводить до множини прийняття гіпотези H_0

$$E_{H_0} = \{ y \in R^{n+1} : \frac{S_{2,0}^2 - S_3^2}{S_3^2} \le \varphi \},$$

де

$$S_{2,0}^2 = \sum_{i=0}^{n} (y_i - \hat\beta_{1,0} x_i - \hat\beta_{2,0} x_i^2)^2,$$

$\hat\beta_{1,0}, \hat\beta_{2,0}$– оцінки МНК невідомих параметрів моделі лінійної регресії

$$y_i = \beta_{1,0} x_i + \beta_{2,0} x_i^2 + \epsilon_i, \qquad i = 0, 1, ..., n.$$

Після низки перетворень маємо

$$S_{2,0}^2 = \sum_{i=0}^{n} y_i^2 - \frac{s_1^2(y) \sum_{i=0}^{n} x_i^4 + 2 s_1(y) s_2(y) \sum_{i=0}^{n} x_i^3 + s_2^2(y) \sum_{i=0}^{n} x_i^4}{\sum_{i=0}^{n} x_i^2 \sum_{i=0}^{n} x_i^4 - (\sum_{i=0}^{n} x_i^3)^2},$$

де

$$s_1(y) = \sum_{i=0}^{n} y_i x_i, \qquad s_2(y) = \sum_{i=0}^{n} y_i x_i^2.$$

Різницю $S_{2,0}^2 - S_3^2$ можна подати у вигляді [2]

$$S_{2,0}^2 - S_3^2 = \frac{\hat{\beta}_0^2}{z_0};$$

у випадку $x_i = \frac{i}{n}, i = 0, 1, ..., n,$

$$z_0 = \frac{3(3n^2 + 3n + 2)}{(n+1)(n+2)(n+3)},$$

а з $(5')$ маємо

$$\hat{\beta}_0 = z_0 L_0(y_0, y_1, ..., y_n).$$

Множина прийняття гіпотези H_0 набирає вигляду

$$E_{H_0} = \Big\{ y \in R^{n+1} : L_0^2(y_0, y_1, ..., y_n) \leq \frac{S_3^2 \varphi}{z_0} \Big\}.$$

3. Умова рівності нулю старшого коефіцієнту поліноміальної регресії

Розглянемо модель поліноміальної регресії

$$y_i = \beta_0 + \beta_1 x_i + ... + \beta_k x_i^k + \epsilon_i, \qquad i = 0, 1, ..., n,$$

де $\beta_0, \beta_1, ..., \beta_k$ — невідомі параметри моделі, які підлягають оцінюванню.

Для знаходження оцінок МНК $\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k$ параметрів $\beta_0, \beta_1, ..., \beta_k$ маємо наступну систему лінійних алгебраїчних рівнянь

$$\begin{cases} \sum_{i=0}^n \big(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i - ... - \hat{\beta}_k x_i^k\big) = 0, \\ \sum_{i=0}^n \big(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i - ... - \hat{\beta}_k x_i^k\big) x_i = 0, \\ \qquad ... \\ \sum_{i=0}^n \big(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i - \hat{\beta}_k x_i^k\big) x_i^k = 0, \end{cases}$$

з якої отримуємо

$$\hat{\beta}_k = \frac{\Delta_k}{\Delta^{(k)}},$$

де

$$\Delta_k = \begin{vmatrix} n+1 & \sum_{j=0}^n x_j & ... & \sum_{j=0}^n x_j^{k-1} & \sum_{i=0}^n y_i \\ \sum_{j=0}^n x_i & \sum_{j=0}^n x_j^2 & ... & \sum_{j=0}^n x_j^k & \sum_{i=0}^n y_i x_i \\ ... & ... & ... & ... & ... \\ \sum_{j=0}^n x_j^k & \sum_{j=0}^n x_j^{k+1} & ... & \sum_{j=0}^n x_j^{2k-1} & \sum_{i=0}^n y_i x_i^k \end{vmatrix},$$

$$\Delta^{(k)} = \begin{vmatrix} n+1 & \sum_{j=0}^n x_j & ... & \sum_{j=0}^n x_j^k \\ \sum_{j=0}^n x_j & \sum_{j=0}^n x_j^2 & ... & \sum_{j=0}^n x_j^{k+1} \\ ... & ... & ... & ... \\ \sum_{j=0}^n x_j^k & \sum_{j=0}^n x_j^{k+1} & ... & \sum_{j=0}^n x_j^{2k} \end{vmatrix}.$$

Очевидно,

$$\hat{\beta}_k = 0 \ \Leftrightarrow \ L_k(y_0, y_1, ..., y_n) = \sum_{i=0}^n a_i^{(k)} y_i = 0,$$

де

$$a_i^{(k)} = \begin{vmatrix} n+1 & \sum_{j=0}^n x_j & ... & \sum_{j=0}^n x_j^{k-1} & 1 \\ \sum_{j=0}^n x_j & \sum_{j=0}^n x_j^2 & ... & \sum_{j=0}^n x_j^k & x_i \\ ... & ... & ... & ... & ... \\ \sum_{j=0}^n x_j^k & \sum_{j=0}^n x_j^{k+1} & ... & \sum_{j=0}^n x_j^{2k-1} & x_i^k \end{vmatrix} = \qquad (10)$$

$$= \delta\Big(x_i^k + \sum_{l=0}^{k-1} \alpha_l^{(k)} x_i^l\Big),$$

δ та $\alpha_l^{(k)}$, $l = 0, 1, ..., k-1$ — визначник та розв'язок лінійної системи

$$\begin{pmatrix} n+1 & \sum_{j=0}^n x_j & ... & \sum_{j=0}^n x_j^{k-1} \\ \sum_{j=0}^n x_j & \sum_{j=0}^n x_j^2 & ... & \sum_{j=0}^n x_j^k \\ ... & ... & ... & ... \\ \sum_{j=0}^n x_j^{k-1} & \sum_{j=0}^n x_j^k & ... & \sum_{j=0}^n x_j^{2k-2} \end{pmatrix} \begin{pmatrix} \alpha_0^{(k)} \\ \alpha_1^{(k)} \\ ... \\ \alpha_{k-1}^{(k)} \end{pmatrix} = \begin{pmatrix} \sum_{j=0}^n x_j^k \\ \sum_{j=0}^n x_j^{k+1} \\ ... \\ \sum_{j=0}^n x_j^{2k-1} \end{pmatrix}. \quad (11)$$

У випадку $x_i = \frac{i}{n}, i = 0, 1, ..., n$ та $k = 2$ ми отримали $\alpha_0^{(2)} = \frac{n-1}{6n}$, $\alpha_1^{(2)} = -1$.

Позначимо $S_l = \sum_{j=0}^n x_j^l$, якщо $x_j = \frac{j}{n}, j = 0, 1, ..., n$. Тоді систему (11) можна переписати у вигляді

$$\begin{pmatrix} S_0 & S_1 & ... & S_{k-1} \\ S_1 & S_2 & ... & S_k \\ ... & ... & ... & ... \\ S_{k-1} & S_k & ... & S_{2k-2} \end{pmatrix} \begin{pmatrix} \alpha_0^{(k)} \\ \alpha_1^{(k)} \\ ... \\ \alpha_{k-1}^{(k)} \end{pmatrix} = \begin{pmatrix} S_k \\ S_{k-1} \\ ... \\ S_{2k-1} \end{pmatrix}. \qquad (12)$$

Помітимо, що

$$\alpha_{k-1}^{(k)} = \frac{S_{2k-1} - \alpha_0^{(k-1)} S_k - ... - \alpha_{k-2}^{(k-1)} S_{2k-2}}{S_{2k-2} - \alpha_0^{(k-1)} S_{k-1} - ... - \alpha_{k-2}^{(k-1)} S_{2k-3}}, \qquad (13)$$

де $\alpha_l^{(k-1)}$, $l = 0, 1, ..., k-1$ — розв'язок системи

$$\begin{pmatrix} n+1 & S_1 & ... & S_{k-2} \\ S_1 & S_2 & ... & S_{k-1} \\ ... & ... & ... & ... \\ S_{k-2} & S_{k-1} & ... & S_{2k-4} \end{pmatrix} \begin{pmatrix} \alpha_0^{(k-1)} \\ \alpha_1^{(k-1)} \\ ... \\ \alpha_{k-2}^{(k-1)} \end{pmatrix} = \begin{pmatrix} S_{k-1} \\ S_k \\ ... \\ S_{2k-3} \end{pmatrix}, \qquad (14)$$

Далі, перші $k-1$ рівнянь системи (12) подамо у вигляді

$$\begin{pmatrix} n+1 & S_1 & ... & S_{k-2} \\ S_1 & S_2 & ... & S_{k-1} \\ ... & ... & ... & ... \\ S_{k-2} & S_{k-1} & ... & S_{2k-4} \end{pmatrix} \begin{pmatrix} \alpha_0^{(k)} \\ \alpha_1^{(k)} \\ ... \\ \alpha_{k-2}^{(k)} \end{pmatrix} = \begin{pmatrix} S_k - \alpha_{k-1}^{(k)} S_{k-1} \\ S_{k+1} - \alpha_{k-1}^{(k)} S_k \\ ... \\ S_{2k-2} - \alpha_{k-1}^{(k)} S_{2k-3} \end{pmatrix}. \qquad (15)$$

Позначимо $\alpha_{l,m}^{(k-1)},\ l = 0, 1, ..., k-1$ — розв'язок лінійної системи

$$\begin{pmatrix} n+1 & S_1 & ... & S_{k-2} \\ S_1 & S_2 & ... & S_{k-1} \\ ... & ... & ... & ... \\ S_{k-2} & S_{k-1} & ... & S_{2k-4} \end{pmatrix} \begin{pmatrix} \alpha_{0,m}^{(k-1)} \\ \alpha_{1,m}^{(k-1)} \\ ... \\ \alpha_{k-2,m}^{(k-1)} \end{pmatrix} = \begin{pmatrix} S_{k-1+m} \\ S_{k+m} \\ ... \\ S_{2k-3+m} \end{pmatrix}, \qquad (16)$$

$$m = 1, 2, ..., k-1.$$

З (15) та (16) отримуємо

$$\alpha_l^{(k)} = \alpha_{l,1}^{(k-1)} - \alpha_{k-1}^{(k)} \alpha_l^{(k-1)},\ l = 0, 1, ..., k-2.$$

Отже, бачимо, що розв'язок системи (13) можна виразити через розв'язок системи (14).

Формули для $S_l,\ l = 1, 2, ...$ знаходимо з формул сум степенів натуральних чисел [4]. Маємо

$$S_1 = \frac{1}{n} \sum_{j=0}^{n} j = \frac{n+1}{2},$$

$$S_2 = \frac{1}{n^2} \sum_{j=0}^{n} j^2 = \frac{(n+1)(2n+1)}{6n},$$

$$S_3 = \frac{1}{n^3} \sum_{j=0}^{n} j^3 = \frac{(n+1)^2}{4n},$$

$$S_4 = \frac{1}{n^4} \sum_{j=0}^{n} j^4 = \frac{(n+1)(2n+1)(3n^2+3n-1)}{30n^3},$$

$$S_5 = \frac{1}{n^5} \sum_{j=0}^{n} j^5 = \frac{(n+1)^2(2n^2+2n-1)}{12n^3},$$

$$S_6 = \frac{1}{n^6} \sum_{j=0}^{n} j^6 = \frac{(n+1)(2n+1)(3n^4+6n^3-3n+1)}{42n^5}.$$

Позначимо

$$\tilde{a}_i^{(k)} = \delta^{-1} a_i^{(k)}, \ i = 0, 1, ..., n. \tag{17}$$

Алгоритм отримання коефіцієнтів $\tilde{a}_i^{(k)}, i = 0, 1, ..., n$, для заданого k.

Знайдемо спочатку $\alpha_l^{(k)}, \ l = 0, 1, ..., k-1$.
1. Знаходимо

$$\alpha_0^{(1)} = \frac{S_1}{S_0}, \quad \alpha_{0,m}^{(1)} = \frac{S_{m+1}}{S_0}, \ m = 1, 2, ..., k-1.$$

2. Знаходимо

$$\alpha_1^{(2)} = \frac{S_3 - \alpha_0^{(1)} S_2}{S_2 - \alpha_0^{(1)} S_1}, \quad \alpha_{1,m}^{(2)} = \frac{S_{3+m} - \alpha_0^{(1)} S_{2+m}}{S_2 - \alpha_0^{(1)} S_1}, \ m = 1, 2, ..., k-2.$$

Далі,

$$\alpha_0^{(2)} = \alpha_{0,1}^{(1)} - \alpha_1^{(2)} \alpha_0^{(1)},$$

$$\alpha_{0,m}^{(2)} = \alpha_{0,m+1}^{(1)} - \alpha_{1,m}^{(2)} \alpha_0^{(1)}, \ m = 1, 2, ..., k-2.$$

3. Для $t = 3, 4, ..., k-1$ знаходимо

$$\alpha_{t-1}^{(t)} = \frac{S_{2t-1} - \alpha_{t-2}^{(t-1)} S_{2t-2} - ... - \alpha_0^{(t-1)} S_t}{S_{2t-2} - \alpha_{t-2}^{(t-1)} S_{2t-3} - ... - \alpha_0^{(t-1)} S_{t-1}},$$

$$\alpha_{t-1,m}^{(t)} = \frac{S_{2t-1+m} - \alpha_{t-2}^{(t-1)} S_{2t-2+m} - ... - \alpha_0^{(t-1)} S_{t+m}}{S_{2t-2} - \alpha_{t-2}^{(t-1)} S_{2t-3} - ... - \alpha_0^{(t-1)} S_{t-1}}, \ m = 1, 2, ..., k-t.$$

Далі,

$$\alpha_{t-2}^{(t)} = \alpha_{t-2,1}^{(t-1)} - \alpha_{t-1}^{(t)} \alpha_{t-2}^{(t-1)},$$

$$\alpha_{t-2,m}^{(t)} = \alpha_{t-2,m+1}^{(t-1)} - \alpha_{t-1,m}^{(t)} \alpha_{t-2}^{(t-1)}, \ m = 1, 2, ..., k-t,$$

$$...$$

$$\alpha_0^{(t)} = \alpha_{0,1}^{(t-1)} - \alpha_{t-1}^{(t)} \alpha_0^{(t-1)},$$

$$\alpha_{0,m}^{(t)} = \alpha_{0,m+1}^{(t-1)} - \alpha_{t-1,m}^{(t)} \alpha_0^{(t-1)}, \ m = 1, 2, ..., k-t.$$

4. Знаходимо $\alpha_{k-1}^{(k)}$ за формулою (13).
Далі,

$$\alpha_{k-2}^{(k)} = \alpha_{k-2,1}^{(k-1)} - \alpha_{k-1}^{(k)}\alpha_{k-2}^{(k-1)},$$

$$\alpha_{k-3}^{(k)} = \alpha_{k-3,1}^{(k-1)} - \alpha_{k-1}^{(k)}\alpha_{k-3}^{(k-1)},$$

$$...$$

$$\alpha_{0}^{(k)} = \alpha_{0,1}^{(k-1)} - \alpha_{k-1}^{(k)}\alpha_{0}^{(k-1)}.$$

5. За допомогою формул (10) та (17) знаходимо $\tilde{a}_i^{(k)},\ i = 0, 1, ..., n$.

Знайдемо $\tilde{a}_i^{(k)},\ i = 0, 1, ..., n$ для $k = 3$ за допомогою цього алгоритма.

$$\alpha_0^{(1)} = \frac{1}{2}, \qquad \alpha_{0,1}^{(1)} = \frac{1}{2} - \frac{n-1}{6n}, \qquad \alpha_{0,2}^{(1)} = \frac{1}{2} - \frac{n-1}{4n};$$

$$\alpha_1^{(2)} = 1, \qquad \alpha_{1,1}^{(2)} = 1 - \frac{(n-1)(n-2)}{10n^2},$$

$$\alpha_0^{(2)} = -\frac{n-1}{6n}, \qquad \alpha_{0,1}^{(2)} = -\frac{(n-1)(2n+1)}{10n^2};$$

$$\alpha_2^{(3)} = \frac{3}{2}, \qquad \alpha_1^{(3)} = -\frac{6n^2-3n+2}{10n^2}, \qquad \alpha_0^{(3)} = \frac{(n-1)(n-2)}{20n^2}.$$

Отже, маємо

$$\tilde{a}_i^{(3)} = x_i^3 - \frac{3}{2}x_i^2 + \frac{6n^2-3n+2}{10n^2}x_i - \frac{(n-1)(n-2)}{20n^2} =$$

$$= (x_i - \frac{1}{2})^3 - \frac{3n^2+6n-4}{20n^2}(x_i - \frac{1}{2}),\ i = 0, 1, ..., n. \tag{18}$$

За допомогою формули (18) отримуємо

$n = 4: \qquad \hat{\beta}_3 = 0 \ \Leftrightarrow\ -y_0 + y_1 - y_3 + y_4 = 0,$

$n = 5: \qquad \hat{\beta}_3 = 0 \ \Leftrightarrow\ -5y_0 + 7y_1 + 4y_2 - 4y_3 - 7y_4 + 5y_5 = 0,$

$n = 6: \qquad \hat{\beta}_3 = 0 \ \Leftrightarrow\ -y_0 + y_1 + y_2 - y_4 - y_5 + y_6 = 0.$

Знайдемо $\tilde{a}_i^{(k)},\ i = 0, 1, ..., n$ для $k = 4$ за допомогою цього алгоритма. Додатково знаходимо

$$\alpha_{0,3}^{(1)} = \frac{(2n+1)(3n^2+3n-1)}{30n^3};$$

$$\alpha_{1,2}^{(2)} = 1 - \frac{(n-1)(n-2)}{5n^2}, \qquad \alpha_{0,2}^{(2)} = -\frac{(n-1)(6n^2+6n-1)}{30n^2};$$

$$\alpha_{2,1}^{(3)} = \frac{12n^2+3n-5}{7n^2}, \qquad \alpha_{1,1}^{(3)} = -\frac{32n^2-6n-11}{35n^2},$$

$$\alpha_{0,1}^{(3)} = \frac{(n-1)(6n^2-9n-6)}{70n^2};$$

$$\alpha_3^{(4)} = 2, \quad \alpha_2^{(4)} = -\frac{9n^2-3n+5}{7n^2}, \quad \alpha_3^{(4)} = \frac{2n^2-3n+5}{7n^2},$$
$$\alpha_0^{(4)} = \frac{(n-1)(n-2)(n-3)}{70n^3}.$$

Отже, маємо

$$\tilde{a}_i^{(4)} = x_i^3 - 2x_i^3 + \frac{9n^2-3n+5}{7n^2}x_i^2 - \frac{2n^2-3n+5}{7n^2}x_i + \frac{(n-1)(n-2)(n-3)}{70n^3} =$$
$$= (x_i - \frac{1}{2})^4 - \frac{1,5n^2+3n-5}{7n^2}(x_i - \frac{1}{2})^2 + \frac{3(n-2)(n+2)(n+4)}{560n^3}, \; i = 0, 1, ..., n. \tag{19}$$

За допомогою формули (19) отримуємо

$$n = 5: \quad \hat{\beta}_4 = 0 \;\Leftrightarrow\; y_0 - 3y_1 + 2y_2 + 2y_3 - 3y_4 + y_5 = 0,$$
$$n = 6: \quad \hat{\beta}_4 = 0 \;\Leftrightarrow\; 3y_0 - 7y_1 + y_2 + 6y_3 + y_4 - 7y_5 + 3y_6 = 0,$$
$$n = 7: \quad \hat{\beta}_4 = 0 \;\Leftrightarrow\; 7y_0 - 13y_1 - 3y_2 + 9y_3 + 9y_4 - 3y_5 - 13y_6 + 7y_7 = 0.$$

Література

1. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. — Москва: Финансы и статистика, 1986. — 366 с.
2. Демиденко Е. З. Линейная и нелинейная регрессии. — Москва: Финансы и статистика, 1981. — 304 с.
3. Себер Дж. Линейный регрессионный анализ. — Москва: Мир, 1980. — 456 с.
4. Абрамович В. С. Суммы одинаковых степеней натуральных чисел // Журнал „Квант“. — 1973. — №5. — С. 22–25.
5. Протасов Ю. М., Юров В. М. Гармонический анализ периодических колебаний объемов продаж компании на основе инструмента „регрессия“ MS EXCEL// Вестник Московского гос. университета, серия: экономика. — 2016. — №2. — С. 115–121 — DOI: 10.18384/2310-6646-2016-2-115-121.
6. Емцева Е. Д., Мазелис А. Л. Моделирование взаимосвязи валового регионального продукта и показателей качества жизни // Вектор науки ТГУ, серия: экономика. — 2016. — №3(26). — С. 24–28 — DOI: 10.18323/2221-5689-2016-3-24-28.
7. Ning Hao, Yang Feng, Hao Helen Zhang Model Selection for High Dimensional Quadratic Regression via Regularization // JASA. — 2016. — Vol. 111, №515 — DOI: 10.1080/01621459.2016.1264956.

Надійшла 15.09.16