УДК 519.2

Савкіна М.Ю., к. ф.-м. н., с.н.с.

Алгоритм перевірки на коректність моделі двофазної нелінійної регресії

Інститут математики НАНУ, 01601, м.Київ, вул.Терещенківська, 3.
E-mail:marta@imath.kiev.ua

M.Yu.Savkina, PhD., Senior Scientific Researcher

The algorithm of checking for correctness of two-phase nonlinear regression model

Institute of mathematics of NASU, 01601, Kyiv, Tereschenkovskaya, 3.
E-mail:marta@imath.kiev.ua

Отримано в явному вигляді формули для оцінок параметрів та залишкової суми квадратів моделі двофазної регресії з відомою точкою перемикання. Побудовано алгоритм перевірки на коректність моделі двофазної регресії з невідомою точкою перемикання.

Ключові слова: метод найменших квадратів, регресійна модель, точка перемикання

Formulas for parameter estimates and residual sum of squares for two-phase regression model with a known point switch are obtained in explicit form. The parameter estimates of model are presented in a linear combination of values of the response y_i; residual sum of squares is presented in a quadratic form of the values of the response y_i. The coefficients for linear combination and quadratic form are found as a functions of i, k and n, where $n+1$ − quantity of the values of the response, k − number of the knot point in which the switch point is located. The algorithm of checking for correctness of two-phase regression model with unknown point switch is constructed. The algorithm is based on the general principles of statistical hypothesis testing in regression analysis. Testing of hypothesis for equality of the unknown parameter to zero is carried out by using the likelihood ratio criterion, which determines to accept or reject the hypothesis by the ratio of the sum of squared deviations caused this hypothesis to the residual sum of squares of the two-phase regression model with the unknown switch point.

Key Words: least square method, regression model, point switch

Статтю представив д.т.н., проф. Кудін В.І.

1. Вступ

Існує багато задач, які потребують вивчення впливу певних факторів $x_1, x_2, ..., x_k$ на кількісну характеристику y якогось об'єкта або явища. Для розв'язання таких задач використовується регресійний аналіз. Зазвичай $x_1, x_2, ..., x_k$ називають незалежними змінними або регресорами, y − вихідною змінною або відкликом. Нехай між незалежними та вихідною змінними існує лінійний зв'язок. Регресійний аналіз базується на припущенні, що із-за впливу різних неврахованих факторів та випадковостей залежність y від $x_1, x_2, ..., x_k$ буде більш або менш відрізнятися від лінійної, тобто її доречно подати в вигляді

$$y = a_1x_1 + a_2x_2 + ... + a_kx_k + \epsilon, \qquad (1)$$

де ϵ −випадкове відхилення. Таким чином маємо лінійну регресійну модель (1). Величини $a_1, a_2, ..., a_k$ називають параметрами моделі; вони невідомі та завдяки наявності ϵ в рівнянні регресії не можуть бути знайдені точно. Для їх визначення застосовуються статистичні методи досідження невідомих параметрів: точкове та інтервальне оцінювання.

Для того, щоб отримати точкові оцінки параметрів $a_1, a_2, ..., a_k$, слід призвести реєстрацію n значень y та кожного фактору x_j, $j = 1, 2, ..., k$, одночасно, тобто при конкретних значеннях факторів $x_{i1}, x_{i2}, ..., x_{ik}$ отримати значень y_i, $i = 1, 2, ..., n$, $n > k$.

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

$$\sum_{i=0}^{n}(y_i - a_1x_{i1} - a_2x_{i2} - ... - a_kx_{ik})^2$$

відносно $a_1, a_2, ..., a_k$.

Позначимо

$$X = \begin{pmatrix} x_{11} & x_{12} & ... & x_{1k} \\ x_{21} & r_{22} & ... & r_{2k} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ x_{n1} & x_{n2} & ... & x_{nk} \end{pmatrix}.$$

Якщо $rang\ X = k$, то оцінка МНК $\hat{a}_1, \hat{a}_2, ..., \hat{a}_k$ параметрів $a_1, a_2, ..., a_k$ єдина, та її можна знайти за формулою

$$\vec{a}^{(\wedge)} = (X^TX)^{-1}X^T\vec{y}, \tag{2}$$

де

$$\vec{a}^{(\wedge)} = (\hat{a}_1, \hat{a}_2, ..., \hat{a}_k)^T, \quad \vec{y} = (y_1, y_2, ..., y_n)^T.$$

Мінімальне значення суми квадратів відхилень

$$S_k^2 = \sum_{i=0}^{n}(y_i - \hat{a}_1x_{i1} - \hat{a}_2x_{i2} - ... - \hat{a}_kx_{ik})^2 \tag{3}$$

називається залишковою сумою квадратів.

Класичний регресійний аналіз заснований на тому, що вигляд моделі (1) відомий з точністю до параметрів, тобто набір незалежних змінних (факторів) заданий однозначно, всі істотні змінні присутні та ніяких альтернативних способів вибору факторів немає. Насправді вибір регресорів, тісно пов'язаний з вибором моделі об'єкта - одна з найскладніших проблем. Поява ЕОМ в середині минулого сторіччя 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]. Крім того, раніше припускалося, "що вже реалізований експеримент, виконаний за деяким планом. "Таким чином, задача зводилася до вибору найкращої процедури обробки даних. "...Останнім часом отримує розвиток новий підхід, в рамках якого запропоновано одночасно вибирати найкращу тріаду: модель-план-метод оцінювання, що відповідає наскільки можливо розглянутій задачі..."[1]

2. Модель двофазної регресії. Задача перевірки моделі на коректність

Розглянемо модель регресії

$$y_i = at_i + b + c_1(t_i - t^*)_+ + \epsilon_i, \qquad i = 0, 1, ..., n, \tag{4}$$

де $\epsilon_0, ..., \epsilon_n$ – незалежні у сукупності нормально розподілені випадкові величини з $E\epsilon_i = 0$ та $D\epsilon_i = \sigma^2$, а $(t_i - t^*)_+$– зрізана степенева функція [2]. Згідно з [3] точка t^* називається точкою перемикання моделі. Якщо вона відома, модель (4) є лінійною по параметрах a, b, c_1, які підлягають оцінюванню. Якщо t^* невідома, модель стає нелінійною по параметрах, а t^* перетворюється на невідомий параметр моделі, який також треба оцінювати.

Крім того, висуваємо гіпотезу $c_1 = 0$. Якщо вона підтвердиться з великою ймовірністю, фактор $(t - t^*)_+$ видаляємо з регресії, тобто модель (4) перетвориться на таку модель

$$y_i = at_i + b + \epsilon_i, \qquad i = 0, 1, ..., n. \tag{5}$$

Для регресійної моделі (5) формула (2) для оцінок МНК параметрів a, b перетворюється на такі співвідношення [3]

$$\hat{a} = \frac{\sum_{i=0}^{n} y_i(t_i - \bar{t})}{\sum_{i=0}^{n}(t_i - \bar{t})^2}, \quad \hat{b} = \bar{Y} - \hat{a}\bar{t},$$

де

$$\bar{Y} = \sum_{i=0}^{n} y_i, \quad \bar{t} = \sum_{i=0}^{n} t_i,$$

а залишкова сумою квадратів дорівнютеме

$$S_2^2 = \sum_{i=0}^{n}(y_i - \bar{Y})^2 - \hat{a}^2\sum_{i=0}^{n}(t_i - \bar{t})^2.$$

3. Формули для оцінок МНК параметрів моделі двофазної регресії з відомою точкою перемикання

Нехай відомо, що $t^* = t_k$. Позначимо $\hat{a}^{(k)}, \hat{b}^{(k)}, \hat{c}_1^{(k)}$ оцінки МНК параметрів a, b, c_1 моделі (4) в цьому випадку.

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

$$\hat{c}_1^{(k)} = \alpha^{(k)} \sum_{i=0}^{n} \beta_i^{(k)} \gamma_i^{(k)} y_i, \tag{6}$$

$$\hat{a}^{(k)} = 12\delta \sum_{i=0}^{n} \rho_i y_i - \lambda^{(k)} \delta^{(k)} \eta^{(k)} \hat{c}_1^{(k)}, \tag{7}$$

$$\hat{b}^{(k)} = \bar{Y} - 6\delta \sum_{i=0}^{n} \rho_i y_i + \frac{k}{n} \delta^{(k)} \eta^{(k)} \hat{c}_1^{(k)}, \tag{8}$$

де

$$\beta_i^{(k)} = \begin{cases} \frac{1}{k(k+1)}, \text{ якщо } i = 0, 1, ..., k, \\ \frac{1}{(n-k)(n-k+1)}, \text{ якщо } i = k+1, ..., n, \end{cases} \tag{9}$$

$$\gamma_i^{(k)} = \begin{cases} kn - i(2k+n+2), \\ \qquad \text{якщо } i = 0, 1, ..., k, \\ (k-2n-2)n - i(2k-3n-2), \\ \qquad \text{якщо } i = k+1, ..., n, \end{cases} \tag{10}$$

$$\rho_i = \frac{i}{n} - \frac{1}{2}, \quad i = 0, 1, ..., n, \tag{11}$$

$$\alpha^{(k)} = \frac{6n}{(2k+1)n + 2(k^2-1)}, \tag{12}$$

$$\delta = \frac{n}{(n+1)(n+2)}, \tag{13}$$

$$\lambda^{(k)} = \frac{n+2k+2}{n}, \tag{14}$$

$$\eta^{(k)} = \frac{(n-k)(n-k+1)}{n}. \tag{15}$$

Доведення. Позначимо $\tau_{ik} = (t_i - t_k)_+$. Якщо $t^* = t_k$, оцінки МНК $\hat{a}^{(k)}, \hat{b}^{(k)}, \hat{c}_1^{(k)}$ параметрів a, b, c_1 можна знайти як розв'язок системи лінійних алгебраїчних рівнянь

$$\begin{cases} \sum_{i=0}^{n} \left(y_i - \hat{a}^{(k)} t_i - \hat{b}^{(k)} - \hat{c}_1^{(k)} \tau_{ik}\right) = 0 \\ \sum_{i=0}^{n} \left(y_i - \hat{a}^{(k)} t_i - \hat{b}^{(k)} - \hat{c}_1^{(k)} \tau_{ik}\right) t_i = 0 \\ \sum_{i=0}^{n} \left(y_i - \hat{a}^{(k)} t_i - \hat{b}^{(k)} - \hat{c}_1^{(k)} \tau_{ik}\right) \tau_{ik} = 0 \end{cases},$$

з якої отримаємо

$$\hat{c}_1^{(k)} = \frac{\Delta_1}{\Delta},$$

де

$$\Delta_1 = \begin{vmatrix} \sum_{i=0}^{n} \left(t_i^2 - \bar{t}^2\right) & \sum_{i=0}^{n} y_i \left(t_i - \bar{t}\right) \\ \sum_{i=0}^{n} (t_i - \bar{t}) \tau_{ik} & \sum_{i=0}^{n} (y_i - \bar{Y}) \tau_{ik} \end{vmatrix},$$

$$\Delta = \begin{vmatrix} \sum_{i=0}^{n} \left(t_i^2 - \bar{t}^2\right) & \sum_{i=0}^{n} (t_i - \bar{t}) \tau_{ik} \\ \sum_{i=0}^{n} (t_i - \bar{t}) \tau_{ik} & \sum_{i=0}^{n} \tau_{ik}^2 - \frac{1}{n} \left(\sum_{i=0}^{n} \tau_{ik}\right)^2 \end{vmatrix},$$

$$\hat{a}^{(k)} = \left(t_i^2 - \bar{t}^2\right)^{-1} \left[\sum_{i=0}^{n} y_i \left(t_i - \bar{t}\right) - \hat{c}_1^{(k)} \sum_{i=k+1}^{n} (t_i - \bar{t}) \tau_{ik}\right],$$

$$\hat{b}^{(k)} = \bar{Y} - \hat{a}^{(k)} \bar{t} - \frac{\hat{c}_1^{(k)}}{n+1} \sum_{i=k+1}^{n} \tau_{ik}.$$

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

$$\Delta_1 = \begin{vmatrix} \frac{(n+1)(n+2)}{12n} & \sum_{i=0}^{n} y_i \rho_i^{(k)} \\ \frac{(n-k)(n-k+1)(n+2k+2)}{12n^2} & \sum_{i=k}^{n} (y_i - \bar{Y}) \varrho_i^{(k)} \end{vmatrix},$$

де $\varrho_i^{(k)} = \frac{i-k}{n}$, $i = k, ..., n$, а $\rho_i^{(k)}$ визначається за формулою (11),

$$\Delta = (n-k)(n-k+1) \cdot$$

$$\cdot \begin{vmatrix} \frac{(n+1)(n+2)}{12n} & \frac{(n-k)(n-k+1)(n+2k+2)}{12n^2} \\ \frac{n+2k+2}{12n^2} & \frac{(n^2+(2k+3)n-3k^2-k+2)}{12n^2(n+1)} \end{vmatrix}.$$

Після низки перетворень маємо

$$\Delta_1 = \frac{1}{12n^3} \cdot$$

$$\cdot \left(\sum_{i=0}^{k} y_i \left[(n-k)(n-k+1) \left(kn - i(2k+n+2)\right) \right] + \right.$$

$$\left. + \sum_{i=k+1}^{n} y_i \left[k(k+1) \left((k-2n-2)n - i(2k-3n-2) \right) \right] \right);$$

$$\Delta = \frac{2k(k+1)(n-k)(n-k+1)}{(12n^2)^2}.$$

$$\cdot\big((2k+1)n + 2(k^2-1)\big).$$

Таким чином маємо

$$\hat{c}_1^{(k)} = \frac{6n}{(2k+1)n+2(k^2-1)}\cdot$$

$$\cdot\left(\frac{1}{k(k+1)}\sum_{i=0}^{k} y_i\big[kn - i(2k+n+2)\big]+\right.$$

$$+\frac{1}{(n-k)(n-k+1)}\cdot$$

$$\left.\sum_{i=k+1}^{n} y_i\big[(k-2n-2)n - i(2k-3n-2)\big]\right),$$

$$\hat{a}^{(k)} = \frac{12n}{(n+1)(n+2)}\sum_{i=0}^{n} y_i\Big(\frac{i}{n}-\frac{1}{2}\Big)-$$

$$-\frac{(n-k)(n-k+1)(n+2k+2)}{n(n+1)(n+2)}\hat{c}_1^{(k)},$$

$$\hat{b}^{(k)} = \bar{Y} - \frac{6n}{(n+1)(n+2)}\sum_{i=0}^{n} y_i\Big(\frac{i}{n}-\frac{1}{2}\Big)+$$

$$+\frac{k(n-k)(n-k+1)}{n(n+1)(n+2)}\hat{c}_1^{(k)}.$$

В отриманих для $\hat{a}^{(k)}, \hat{b}^{(k)}, \hat{c}_1^{(k)}$ формулах вводимо позначення (9)-(15). Отримаємо формули (6),(7),(8).

Теорему доведено.

Залишкову сумою квадратів для моделі (4) з відомою точкою перемикання можна подати у вигляді

$$T(k) = \sum_{i=0}^{n}\big(y_i - \hat{a}^{(k)}t_i - \hat{b}^{(k)} - \hat{c}_1^{(k)}(t_i - t_k)_+\big)^2. \tag{16}$$

Наслідок. Має місце рівність

$$T(k) = \sum_{i=0}^{n}(y_i - \bar{Y})^2 - 12\delta\Big(\sum_{i=0}^{n}\rho_i y_i\Big)^2+$$

$$+2\hat{c}_1^{(k)}A_1(k) + (\hat{c}_1^{(k)})^2 A_2(k), \tag{17}$$

де

$$A_1(k) = \frac{\eta^{(k)}}{2}\bar{Y} + \eta^{(k)}\lambda^{(k)}\delta\Big(\sum_{i=0}^{n}\rho_i^{(k)}y_i\Big)-$$

$$-\sum_{i=0}^{n}\varrho_i^{(k)}y_i,$$

$$A_2(k) = \frac{(\eta^{(k)})^2}{12(n+1)}\Big((\lambda^{(k)})^2\frac{n}{n+2}+3\Big)+\frac{\eta^{(k)}\mu^{(k)}}{6}.$$

Доведення. Підставимо (6),(7),(8) в (16). Після низки перетворень отримаємо (17).

4. Отримання оцінки МНК точки перемикання в нелінійному випадку

В роботі [4] розроблено алгоритм отримання оцінки МНК $\hat{a}, \hat{b}, \hat{c}_1, \hat{t}^*$ параметрів a, b, c_1, t^* нелінійної моделі (4) за скінченну кількість кроків.

Розглянемо відрізок $[t_k, t_{k+1}]$.

Згідно алгоритму, якщо оцінка МНК $\hat{t}^*$ параметра t^* належить (t_k, t_{k+1}), то її можна знайти як точку перетину двох прямих $a^{(k)}t + b^{(k)}$ та $c^{(k)}t + d^{(k)}$, перша з яких мінімізує суму

$$\sum_{i=0}^{k}(y_i - a^{(k)}t_i - b^{(k)})^2 \text{ відносно } a^{(k)}, b^{(k)},$$

інша - суму

$$\sum_{i=k+1}^{n}(y_i - c^{(k)}t_i - d^{(k)})^2 \text{ відносно } c^{(k)}, d^{(k)}.$$

Якщо прямі $a^{(k)}t+b^{(k)}$ та $c^{(k)}t+d^{(k)}$ не перетинаються на інтервалі (t_k, t_{k+1}), то оцінка МНК $\hat{t}^*$ не належить цьому інтервалу.

1-й крок алгоритма. Для $k = 1, 2, ..., n-2$ будуємо прямі $a^{(k)}t + b^{(k)}$ та $c^{(k)}t + d^{(k)}$ та знаходимо їх точку перетину $\hat{t}_k^*$; якщо вона належить (t_k, t_{k+1}), покладемо

$$S(k) = \sum_{i=0}^{k}\big(y_i - \bar{Y}_1^{(k)}\big)^2 + \sum_{i=k+1}^{n}\big(y_i - \bar{Y}_2^{(k)}\big)^2-$$

$$-\big(a^{(k)}\big)^2\sum_{i=0}^{k}\big(t_i - \bar{t}_1^{(k)}\big)^2 - \big(c^{(k)}\big)^2\sum_{i=k+1}^{n}\big(t_i - \bar{t}_2^{(k)}\big)^2,$$

де

$$\bar{Y}_1^{(k)} = \sum_{i=0}^{k} y_i, \quad \bar{Y}_2^{(k)} = \sum_{i=k+1}^{n} y_i,$$

$$\bar{t}_1^{(k)} = \sum_{i=0}^{k} t_i, \quad \bar{t}_2^{(k)} = \sum_{i=k+1}^{n} t_i;$$

інакше

$$S(k) = \infty.$$

Таким чином, отримаємо послідовність $S(1), S(2), ..., S(n-2)$.

2-й крок алгоритма. Знайдемо в послідовності $S(1), S(2), ..., S(n-2)$ всі пари $\big(S(k_l - 1),\ S(k_l)\big)$, $l = 0, 1, ..., j$, такі що $S(k_l - 1) = \infty$, $S(k_l) = \infty$.

Вважаємо $t^* = t_{k_l}$, $l = 0, 1, ..., j$, та знаходимо $T(k_l)$ за формулою (17); отримаємо другу послідовність $T(k_1), T(k_2), ..., T(k_j)$.

Кожному значенню $S(k) < \infty$ відповідає точка t_k^*, кожному значенню $T(k_l)$ - точка t_{k_l}. Оцінкою МНК $\hat{t}^*$ параметра t^* нелінійної регресійної моделі (3) буде та точка t_k^* або t_{k_l}, яка відповідає значенню

$$S_4^2 = min\{S(1), S(2), ..., S(n-2),$$
$$T(k_1), T(k_2), ..., T(k_j)\}.$$

Далі, в випадку $t_i = \frac{i}{n}, i = 0, 1, ..., n$, маємо

$$\hat{a} = 12\delta \sum_{i=0}^{n} \rho_i y_i, \quad \hat{b} = \bar{Y} - \frac{1}{2}\hat{a}, \tag{18}$$

$$S_2^2 = \sum_{i=0}^{n} (y_i - \bar{Y})^2 - 12\delta \Big(\sum_{i=0}^{n} \rho_i y_i)\Big)^2. \tag{19}$$

$$a^{(k)} = \frac{12n\delta}{k} \sum_{i=0}^{k} y_i \Big(\frac{i}{n} - \frac{k}{2n}\Big), \tag{20}$$

$$c^{(k)} = \frac{12n}{(n-k-1)\eta^{(k)}} \sum_{i=k+1}^{n} y_i \Big(\frac{i}{n} - \frac{n+k+1}{2n}\Big), \tag{21}$$

$$S(k) = \sum_{i=0}^{k} \big(y_i - \bar{Y}_1^{(k)}\big)^2 + \sum_{i=k+1}^{n} \big(y_i - \bar{Y}_2^{(k)}\big)^2 -$$
$$-\frac{12n\delta}{k} \Big(\sum_{i=0}^{k} y_i \Big(\frac{i}{n} - \frac{k}{2n}\Big)\Big)^2 -$$
$$-\frac{12n}{(n-k-1)\eta^{(k)}} \Big(\sum_{i=k+1}^{n} y_i \Big(\frac{i}{n} - \frac{n+k+1}{2n}\Big)\Big)^2. \tag{22}$$

5. Алгоритм прийняття (або відхилення) гіпотези $c_1 = 0$ у випадку невідомої точки перемикання

Розглянемо тепер задачу перевірки простої статистичної гіпотези

$$H:\ c_1 = 0.$$

Критерій відношення правдоподібності перевірки гіпотези H приводить до множини прийняття гіпотези [5]

$$E_H = \{y \in R^{n+1} : \frac{S_2^2 - S_4^2}{S_4^2} \le \varphi\}.$$

Значення φ для нелінійної регресії можна вибирати різними методами. Один з них співпадає з методом вибору в лінійній регресії. Припустимо, модель лінійній регресії має m невідомих параметрів, а гіпотеза H_0 складається з p лінійних рівнянь. Задамося рівнем значущості α та знайдемо для нього відповідне значення $F_\alpha(p, n+1-m)$ наступним чином. Позначимо $f(t, p, n+1-m)-$ 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

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

В нашому випадку $m = 4$ (модель має 4 невідомі параметри), та $p = 1$ (гіпотеза H складається з одного рівняння), тобто

$$\varphi = \frac{1}{n-3} F_\alpha(1, n-3).$$

1) Знаходимо $\hat{a}, \hat{b}, S_2^2$ за формулами (18),(19);
2) Знаходимо $M_1 = \max\{|y_0 - \hat{b}|,\ |y_n - \hat{a} - \hat{b}|\}$; якщо

$$\frac{M_1^2}{S_2^2 - M_1^2} \ge \varphi, \tag{23}$$

гіпотезу $c_1 = 0$ відхиляємо, інакше переходимо до 3).

3) Знаходимо

$$M_2 = \max\{\left|y_i - \hat{a}\frac{i}{n} - \hat{b}\right|,\ i = 1, ..., n-1\}.$$

Нехай $M_2 = \left|y_l - \hat{a}\frac{l}{n} - \hat{b}\right|$, тобто найбільшого значення величина $\left|y_i - \hat{a}\frac{i}{n} - \hat{b}\right|$ досягає при $i = l,\ 1 \leq l \leq n-1$.
Знаходимо $T(l)$ за формулою (14);
якщо

$$\frac{T(l)}{S_2^2 - T(l)} \geq \varphi, \tag{24}$$

гіпотезу $c_1 = 0$ відхиляємо, інакше переходимо до 3).
4) Для $k = 1, ..., n-1:$
знаходимо $\hat{a}^{(k)}, \hat{b}^{(k)}$ за формулами (8),(9); якщо виконується нерівність

$$\left|(c^{(k)} - a^{(k)})^{-1}\left(\bar{Y}_1^{(k)} - \bar{Y}_2^{(k)} - \frac{c^{(k)}}{2}\left(1 + \frac{1}{n}\right)\right) - \right.$$

$$\left. - \frac{k}{2n}\right| < \frac{1}{n},$$

знаходимо $S(k)$ за формулою (10), інакше покладемо $S(k) = \infty$.
5)Знаходимо $T(k_1), T(k_2), ..., T(k_j)\}$ за формулою (14).
6)Знаходимо S_4^2. Якщо

$$\frac{S_2^2 - S_4^2}{S_4^2} \geq \varphi, \tag{25}$$

гіпотезу $c_1 = 0$ відхиляємо, інакше приймаємо.

З а у в а ж е н н я. З нерівностей (23),(24) відразу випливає (25), тому що

$$M_1^2 < S_2^2 - S_4^2,\ T(l) < S_2^2 - S_4^2.$$

Якщо нерівності (23),(24) не виконуються, найвірогідніше гіпотеза $c_1 = 0$ буде прийнята. Прийняття цієї гіпотези може відбутися тільки на кроці 6); відхилення гіпотези скоріше за все станеться на кроці 2) або 3), тому в цьому випадку шукати S_4^2 не треба.

Список використаних джерел

1. *Дрейпер Н.* Прикладной регрессионный анализ / Н. Дрейпер, Г. Смит.— Москва: Финансы и статистика, 1986.—366 с.
2. *Завьялов Ю.С.* Методы сплайн-функций / Ю.С. Завьялов, Б.И. Квасов, В.Л. Мирошниченко. — Москва: Наука, 1980.—352 с.
3. *Себер Дж.* Линейный регрессионный анализ / Дж. Себер. — Москва: Мир, 1980.—456 с.
4. *Hudson Derek J.* Fitting Segmented Curves Whose Join Points Have to Be Estimated / Derek J.Hudson // JASA. — 1966. — v.61. — N.316. — P.1097–1129.
5. *Демиденко Е.З.* Линейная и нелинейная регрессии / Е.З. Демиденко. — Москва: Финансы и статистика, 1981.—304 с.

References

1. DRAPER N., SMITH H.(1986) Applied regression analisis. Moscow: Finansy and statistika.
2. ZAV'YALOV Y.S., KVASOV B., MIROSHNICHENKO V.(1980). Methods of spline functions. Moscow: Science.
3. SEBER G.(1980) Linear regression analisis. Moscow: Mir.
4. HUDSON D.(1966) Fitting Segmented Curves Whose Join Points Have to Be Estimated. JASA. 61(316). p.1097–1129.
5. DEMIDENKO E.Z.(1981) Linear and nonlinear regressions. Moscow: Finansy and statistika.

Надійшла до редколегії 18.09.2015.