ОПРАЦЮВАННЯ ТА ПЕРЕТВОРЕННЯ ВИМІРЮВАЛЬНИХ СИГНАЛІВ

УДК 621.317

ДОСЛІДЖЕННЯ ВПЛИВУ КОРЕЛЯЦІЇ РЕЗУЛЬТАТІВ СПОСТЕРЕЖЕНЬ НА НЕПЕВНІСТЬ КОЕФІЦІЄНТІВ ТА ПРОГНОЗОВАНИХ ЗНАЧЕНЬ ЛІНІЙНОЇ РЕГРЕСІЇ

© Дорожовець Михайло, 2008

Національний університет “Львівська політехніка”,
кафедра інформаційно-вимірювальних технологій, вул. С. Бандери, 12, Львів, Україна
dorozhovets@polynet.lviv.ua
Ряшівська політехніка, вул. В.Поля, 2В, 35-959, Ряшів, Польща
michdor@prz.rzeszow.pl

Подано результати досліджень впливу взаємної кореляції результатів спостережень на стандартну непевність коефіцієнтів та прогнозованих значень лінійної регресії. Наведено аналітичні вирази для обчислення стандартних непевностей цих величин залежно від виду автокореляційної функції результатів спостережень.

Представлены результаты исследований влияния взаимной коррелированности результатов наблюдений на стандартную неопределенность коэффициентов и прогнозируемых значений линейной регрессии. Приведены аналитические выражения для вычисления стандартных неопределенностей этих величин в зависимости от вида автокорреляционной функции результатов наблюдений.

In the article the results of studies of the influence of the mutual correlation of the observations to the standard uncertainty of coefficients and forecasted values of linear regression are represented. The analytical expressions for enumerating the standard uncertainties of these values in the dependence form the autocorrelation function of the observations are given.

1. Вступ. Лінійний регресійний аналіз широко використовується для знаходження залежностей між величинами у різних галузях науки, зокрема, у під час біометричних досліджень, а також під час опрацювання результатів технічних вимірювань [1–10]. Під час вимірювань дуже часто результати спостережень супроводжуються систематичними зміщеннями, зокрема у вигляді лінійного часового тренду, який необхідно виявити та скоригувати. Для цього також можна застосувати лінійний регресійний аналіз. Важливим є не лише знаходження скоригованого результату вимірювання, але також і оцінювання його непевності [4, 5, 11–13].

Загалом математична модель лінійної залежності вихідної величини y від вхідної x у заданих експериментально визначених точках (y_i, x_i) (i=1, ..., n, де n – кількість експериментальних точок) описується простим виразом [1–8]

$$y_i = a_0 + a_1 x_i + v_i . \tag{1}$$

де a_0 та a_1 – коефіцієнти залежності, які треба знайти; x_i – значення вхідної величини, що задані точно; y_i – значення вихідної величини, які виміряні не точно; v_i – відхилення результатів вимірювань вхідної величини, що підпорядковані нормальному розподілу із нульовим математичним сподіванням $m_v = M[v_i] = 0$, дисперсією $s_v^2 = M\left[(v_i - m_v)^2\right]$. Загалом під час досліджень приймають, що почергові значення вихідної величини взаємно некорельовані, тобто коефіцієнт кореляції між значеннями v_i та v_j дорівнює нулеві $r_{i,j} = 0$.

Така постановка задачі відповідає задачі лінійної регресії величини y відносно величини x $y(x)$ [6–8], розв'язки якої знаходять, застосовуючи до математичної моделі (1) процедуру методу найменших квадратів МНК, за якої мінімізують суму квадратів відхилень експериментальних точок від шуканої прямої лінії

$$V^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - (a_0 + a_1 x_i)\right)^2 \Rightarrow \min . \qquad (2)$$

Оцінки коефіцієнтів нахилу a_1 та початкового зміщення a_0 лінійної залежності (1) знаходять за виразами [6–9]

$$a_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{S_x^2} = r_{xy}\frac{S_y}{S_x} \text{ (а); } a_0 = \bar{y} - a_1\bar{x} \text{ (б), } (3)$$

де

$$S_x^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2, \; S_y^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 \; - \qquad (4)$$

центральні моменти другого порядку вхідної (S_x^2 невипадкове значення) та вихідної (S_y^2 випадкове значення) величин;

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \; - \qquad (5)$$

– середні значення вхідної (невипадкове) та вихідної (випадкове) величин;

$$r_{xy} = \frac{R_{xy}}{S_x S_y} \; - \qquad (6)$$

експериментальна оцінка коефіцієнта кореляції між вхідною та вихідною величинами

$$R_{xy} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right) = r_{xy}S_x S_y \; - \qquad (7)$$

експериментальна оцінка коваріаційного моменту між вхідною та вихідною величинами.

Оскільки відповідно до (1)

$$y_i - \bar{y} = a_1\left(x_i - \bar{x}\right) + v_i, \qquad (8)$$

то згідно з (3, а) математичне сподівання оцінки коефіцієнта нахилу

$$M[a_1] = M\left[\frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(a_1\left(x_i - \bar{x}\right) + v_i\right)}{S_x^2}\right] =$$

$$= a_1 + \frac{1}{S_x^2}M\left[\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\cdot v_i\right] = a_1, \qquad (9)$$

тобто дорівнює коефіцієнту математичної моделі.

Аналогічно математичне сподівання оцінки коефіцієнта зміщення дорівнює коефіцієнту математичної моделі

$$M[a_0] = M\left[\bar{y} - a_1\bar{x}\right] = M\left[a_0 + \left(a_1 - a_1\right)\bar{x} + \frac{1}{n}\sum_{i=1}^{n} v_i\right] =$$

$$= a_0 + M\left[\frac{1}{n}\sum_{i=1}^{n} v_i\right] = a_0 . \qquad (10)$$

За некорельованих результатів вимірювань вихідної величини ($r_{i,j} = 0, i \neq j$) дисперсії знайдених коефіцієнтів описуються виразами

$$s_{a_0}^2 = M\left[\left(a_0 - a_0\right)^2\right] = \frac{s_v^2}{n}\cdot\left(1 + \frac{\left(\bar{x}\right)^2}{S_x^2}\right) \text{ (а);}$$

$$s_{a_1}^2 = M\left[\left(a_1 - a_1\right)^2\right] = \frac{s_v^2}{n}\cdot\frac{1}{S_x^2} \text{ (б),} \qquad (11)$$

а взаємна коваріація та коефіцієнт кореляції між коефіцієнтами a_0 та a_1 лінійної апроксимації описуються залежностями

$$R_{a_0 a_1} = M\left[\left(a_0 - a_0\right)\left(a_1 - a_1\right)\right] = \frac{\bar{x}}{S_x^2}\frac{s_v^2}{n},$$

$$r_{a_0 a_1} = \frac{\bar{x}}{\sqrt{S_x^2 + \left(\bar{x}\right)^2}} . \qquad (12)$$

Вираз для знаходження дисперсії апроксимуючої залежності (прогнозованих значень вихідної величини має вигляд

$$s_y^2(x) = M\left[\left(a_0 - a_0 + \left(a_1 - a_1\right)x\right)^2\right] = \frac{s_v^2}{n}\left(1 + \frac{\left(x - \bar{x}\right)^2}{S_x^2}\right). \quad (13)$$

Якщо дисперсія s_v^2 відхилень вихідної величини невідома, то знаходять її експериментальну оцінку за виразом:

$$S_v^2 = \frac{\sum_{i=1}^{n}\left|y_i - \left(a_0 + a_1 x_i\right)\right|^2}{n-2} = \frac{n}{n-2}S_y^2\left(1 - r_{xy}^2\right), \quad (14)$$

Тоді експериментальні оцінки дисперсій знайдених коефіцієнтів

$$s_{a_0}^2 \approx \frac{S_v^2}{n}\cdot\left(1 + \frac{\left(\bar{x}\right)^2}{S_x^2}\right) = S_y^2\cdot\frac{1 - r_{xy}^2}{n-2}\cdot\left(1 + \frac{\left(\bar{x}\right)^2}{S_x^2}\right) \text{ (а);}$$

$$s_{a_1}^2 \approx \frac{S_v^2}{n}\cdot\frac{1}{S_x^2} = \frac{S_y^2}{S_x^2}\cdot\frac{1 - r_{xy}^2}{n-2} \text{ (б),} \qquad (15)$$

а вираз для оцінки дисперсії прогнозованих значень вихідної величини

$$s_y^2(x) \approx \frac{S_v^2}{n}\left(1 + \frac{\left(x - \bar{x}\right)^2}{S_x^2}\right) = \frac{S_y^2\left(1 - r_{xy}^2\right)}{n-2}\left(1 + \frac{\left(x - \bar{x}\right)^2}{S_x^2}\right). \quad (16)$$

Насправді під час досліджень лінійних часових трендів результати спостережень можуть виявитися взаємно корельованими. Така кореляція може бути зумовлена, наприклад, занадто великою частотою дискретизації досліджуваної величини, зокрема, з метою покращання точності оцінок коефіцієнтів (15) та прогнозованих значень величини (16) реєстрацією більшої кількості результатів спостережень (збільшення n). Такий шлях може призвести до хибних (занадто оптимістичних) оцінок шуканих величин, що ніяк не відповідає дійсності.

Метою подальших досліджень є встановлення залежностей експериментальних оцінок дисперсій коефіцієнтів лінійної регресії та її прогнозованих значень від корельованості зареєстрованих результатів спостережень.

2. Дослідження впливу кореляції зареєстрованих результатів спостережень. У подальших дослідженнях приймаємо, що відхилення v_i та v_j експериментальних точок (вихідної величини) від прямої лінії можуть бути взаємно корельованими з коефіцієнтом кореляції $r_{i,j} = M\left[(v_i - m_v)(v_i - m_v)\right] / s_v^2$. Якщо дискретизація є рівномірною (значення аргументу задані у рівновіддалених точках: $x_{i+1} = x_i + h$, $h = const$, то $r_{i,j}$ залежить лише від різниці номерів точок $r_{i,j} = r_{i-j}$.

2.1. Зміщення вхідної величин. Для спрощення подальших викладок щодо досліджень впливу кореляції між результатами спостережень вихідної величини змістимо значення вхідної величин на її середнє значення

$$x_i' = x_i - \overline{x}. \quad (17)$$

Очевидно, що $\overline{x'} = \frac{1}{n}\sum_{i=1}^{n} x_i' = 0$. Для наведених замін математична модель лінійної залежності (1) набуває вигляду

$$y_i = a_0 + a_1\left(x_i - \overline{x} + \overline{x}\right) + v_i =$$
$$= a_0 + a_1\overline{x} + b_1 x_i' + v_i = b_0 + a_1 x_i' + v_i. \quad (18)$$

Тобто коефіцієнт нахилу прямої (18) не зміниться, а початкове зміщення набуває нового значення

$$b_0 = a_0 + a_1\overline{x}. \quad (19)$$

У разі заміни (17) значення основних розрахункових величин (4), (6) та (7) не зміняться:

$$S_{x'}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 = S_x^2;$$
$$R_{x'y'} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right) = r_{xy} S_x S_y,$$
$$r_{x'y'} = \frac{R_{x'y'}}{S_{x'} S_{y'}} = r_{xy}. \quad (20)$$

Тому, оскільки вхідна величина є невипадковою, то зміщення аргументу на її середнє значення $\overline{x}$ не зумовлює жодних змін дисперсії ні оцінки коефіцієнта нахилу прямої, ні оцінки дисперсії прогнозованих значень функції

$$s_y^2(x) = s_y^2(x'). \quad (21)$$

А внаслідок зміщень вхідної величини дисперсія оцінки коефіцієнта початкового зміщення прямої змінюється масштабно на невипадкове значення $1 + \left(\overline{x}\right)^2 / S_x^2$, тобто

$$s_{a_0}^2 = s_{b_0}^2 \cdot \left(1 + \frac{\left(\overline{x}\right)^2}{S_x^2}\right). \quad (22)$$

2.12. Оцінки дисперсій коефіцієнтів лінії. За корельованих результатів вимірювань вихідної величини ($r_{i,j} = r_{i-j} \neq 0, i \neq j$) згідно з (9) та (10) математичні сподівання коефіцієнтів лінії b_0, a_1 також дорівнюють значенням коефіцієнтів математичної моделі. Натомість дисперсії знайдених коефіцієнтів та прогнозованих значень внаслідок взаємної кореляції результатів вимірювань будуть істотно відрізнятися від отриманих вище (11), (13) та (15) і (16). Дослідимо їх.

Якщо $r_{i,j} = r_{i-j} \neq 0,\ i \neq j$, то з урахуванням (10) дисперсія коефіцієнта зміщення описується залежністю [14]

$$s_{b_0,kor}^2 = M\left[\left(b_0 - m_{b_0}\right)^2\right] = M\left[\left(b_0 - b_0\right)^2\right] =$$
$$= M\left[\left(\frac{1}{n}\sum_{i=1}^{n} v_i\right)^2\right] = \frac{s_v^2}{n}\left(1 + \frac{2}{n}\sum_{k=1}^{n-1}(n-k) r_k\right). \quad (23)$$

Аналогічно у разі корельованих результатів вимірювань дисперсія коефіцієнта нахилу з урахуванням (10) та (3, а) описується залежністю

$$s_{a_1,kor}^2 = M\left[\left(a_1 - a_1\right)^2\right] = M\left[\left(\frac{1}{S_x^2}\frac{1}{n}\sum_{i=1}^{n} x_i' v_i\right)^2\right] =$$
$$= \frac{s_v^2}{S_x^2 n}\left(1 + \frac{2}{n}\sum_{k=1}^{n-1} r_k \sum_{j=1}^{n-k} \frac{x_j' x_{k+j}'}{S_x^2}\right). \quad (24)$$

Якщо дисперсія s_v^2 відхилень вихідної величини невідома, то внаслідок взаємної кореляції її експериментальна оцінка також є іншою, порівняно з отриманою вище у (14). Середнє значення суми квадратів відхилень (2) після підстановки виразів (3) з урахуванням (17) набуває вигляду

$$S_v^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \left(b_0 + a_1 x_i'\right)\right)^2 =$$
$$= \frac{1}{n}\sum_{i=1}^{n}\left(v_i - \bar{v} - \frac{x_i'}{nS_x^2}\sum_{j=1}^{n} x_j' v_j\right)^2 . \qquad (25)$$

Математичне сподівання цього значення знайдемо почленно, розкриваючи дужки і виконуючи з отриманими складовими операції математичного сподівання:

$$M\left[S_v^2\right]_{kor} = M\left[\frac{1}{n}\sum_{i=1}^{n} v_i^2\right] +$$
$$+M\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{n}\sum_{j=1}^{n} v_j\right] + \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i'}{nS_x^2}\sum_{j=1}^{n} x_j' v_j\right)^2 -$$
$$-2M\left[\frac{1}{n}\sum_{i=1}^{n} v_i \frac{1}{n}\sum_{j=1}^{n} v_j\right] - \qquad (26)$$
$$-2M\left[\frac{1}{n}\sum_{j=1}^{n}\frac{v_i x_i'}{nS_x^2}\sum_{j=1}^{n} x_j v_j\right] +$$
$$+2M\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{n}\sum_{j=1}^{n} v_j\left(\frac{x_i'}{nS_x^2}\sum_{j=1}^{n} x_j' v_j\right)\right] .$$

Тоді значення математичних сподівань складових у (26) становлять:

$$M\left[\frac{1}{n}\sum_{i=1}^{n} v_i^2\right] = s_v^2 . \qquad (27)$$

$$M\left[\left(\frac{1}{n}\sum_{i=1}^{n} v_i\right)^2\right] = \frac{s_v^2}{n}\left(1 + \frac{2}{n}\sum_{k=1}^{n-1}(n-k) r_k\right). \qquad (28)$$

$$M\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i'}{nS_x^2}\sum_{j=1}^{n} x_j' v_j\right)^2\right] = \frac{s_v^2}{n}\left(1 + \frac{2}{n}\sum_{k=1}^{n-1} r_k \sum_{j=1}^{n-k}\frac{x_j' x_{k+j}'}{S_x^2}\right); (29)$$

$$M\left[\frac{1}{n}\sum_{i=1}^{n} v_i \frac{1}{n}\sum_{j=1}^{n} v_j\right] = \frac{s_v^2}{n}\left(1 + \frac{2}{n}\sum_{k=1}^{n-1}(n-k) r_k\right); \quad (30)$$

$$M\left[\frac{1}{n}\sum_{j=1}^{n}\frac{v_i x_i'}{nS_x^2}\sum_{j=1}^{n} x_j' v_j\right] = \frac{s_v^2}{n}\left(1 + \frac{2}{n}\sum_{k=1}^{n-1} r_k \sum_{j=1}^{n-k}\frac{x_j' x_{k+j}'}{S_x^2}\right); \; 31)$$

і, нарешті, оскільки $\overline{x'} = 0$

$$M\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{n}\sum_{j=1}^{n} v_j\left(\frac{x_i'}{nS_x^2}\sum_{j=1}^{n} x_j' v_j\right)\right] = 0 . \qquad (32)$$

Після підстановки (27) – (32) у (26) отримуємо математичне сподівання експериментального середнього значення суми квадратів відхилень (25)

$$M\left[S_v^2\right]_{kor} = s_v^2\left[1 - \frac{1}{n}\left(1 + \frac{2}{n}\sum_{k=1}^{n-1}(n-k) r_k\right) - \right.$$
$$\left. - \frac{1}{n}\left(1 + \frac{2}{n}\sum_{k=1}^{n-1} r_k \sum_{j=1}^{n-k}\frac{x_j' x_{k+j}'}{S_x^2}\right)\right] \qquad (33)$$

Підстановкою $r_k = 0, k \neq 0$ у (33) встановлюємо, що $M\left[S_v^2\right]_{r=0} = s_v^2\left(1 - \frac{2}{n}\right)$, що є теоретичною оцінкою для некорельованих результатів спостережень.

Із (33) випливає, що оцінка експериментального середнього значення суми квадратів відхилень (25) є зміщеною і для усунення цього зміщення (25) необхідно поділити на вираз у квадратних дужках (33). Тобто

$$S_{v,н.зм.}^2 = \frac{S_v^2}{1 - \frac{1}{n}\left(1 + \frac{2}{n}\sum_{k=1}^{n-1}(n-k) r_k\right) - \frac{1}{n}\left(1 + \frac{2}{n}\sum_{k=1}^{n-1} r_k \sum_{j=1}^{n-k}\frac{x_j' x_{k+j}'}{S_x^2}\right)} . (34)$$

У знаменнику (34) зробимо перетворення

$$1 - \frac{1}{n}\left(1 + \frac{2}{n}\sum_{k=1}^{n-1}(n-k) r_k\right) -$$
$$- \frac{1}{n}\left(1 + \frac{2}{n}\sum_{k=1}^{n-1} r_k \sum_{j=1}^{n-k}\frac{x_j' x_{k+j}'}{S_x^2}\right) =$$
$$= 1 - \frac{2}{n}\left(1 + \frac{2}{n}\sum_{k=1}^{n-1}(n-k) r_k\right) + \qquad (35)$$
$$+ \frac{2}{n^2}\left(\sum_{k=1}^{n-1} r_k\left[(n-k) - \sum_{j=1}^{n-k}\frac{x_j' x_{k+j}'}{S_x^2}\right]\right) .$$

Введемо позначення ефективної кількості результатів спостережень першого роду [14]

$$n_{eff.1} = \frac{n}{1 + \frac{2}{n}\sum_{k=1}^{n-1}(n-k) r_k} , \qquad (36)$$

яка може набувати значень від $n_{eff.1} = n$ при некорельованих результатах спостережень $r_k = 0, k \neq 0$ до $n_{eff.1} = 1$ при повністю корельованих результатах $r_k = 1$, що є цілком логічним.

А також використаємо позначення ефективної кількості результатів спостережень другого роду

$$n_{eff.2}=\frac{n}{\frac{2}{n}\sum_{k=1}^{n-1}r_k\left[(n-k)-\sum_{j=1}^{n-k}\frac{x'_j x'_{k+j}}{S_x^2}\right]} \qquad (37)$$

яка набуває нескінченного значення $n_{eff.2}=\infty$ при некорельованих результатах спостережень $r_k=0, k\neq 0$ і $n_{eff.2}=1$ при повністю корельованих результатах $r_k=1$.

У разі рівновіддалених значень вхідної величини з інтервалом $h=\frac{x_B-x_A}{n-1}$, де x_A, x_B – початок та кінець інтервалу, тобто $x_i=x_A+h\cdot(i-1), i=1,2,...n$, середнє значення вхідної величини становить: $\bar{x}=(x_B-x_A)/2$ і центровані значення вхідної величини описуються залежністю

$$x'_i=x_i-\bar{x}=x_A+h\cdot(i-1)-\frac{x_B+x_A}{2}=h\cdot\left(i-\frac{n+1}{2}\right). \qquad (38)$$

Використовуючи (38) у (4), знаходимо середнє значення квадратів зміщених значень вхідної величини

$$S_x^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2=\frac{n^2-1}{12}h^2. \qquad (39)$$

Підставляючи значення (38) і (39) у вираз $\sum_{j=1}^{n-k}\frac{x'_j x'_{k+j}}{S_x^2}$ (37), знаходимо

$$\sum_{j=1}^{n-k}\frac{x'_j x'_{k+j}}{S_x^2}=\frac{h^2}{\frac{n^2-1}{12}h^2}\sum_{j=1}^{n-k}\left(j-\frac{n+1}{2}\right)\cdot\left(k+j-\frac{n+1}{2}\right)=$$

$$=(n-k)\left(1-\frac{2k(n+k)}{n^2-1}\right). \qquad (40)$$

Отже, ефективна кількість результатів спостережень другого роду у (37) описується спрощеною залежністю

$$n_{eff.2}=\frac{n}{\frac{2}{n^3}\sum_{k=1}^{n-1}r_k\left[(n-k)-(n-k)\left(1-\frac{2k(n+k)}{n^2-1}\right)\right]}=$$

$$=\frac{n}{\frac{4}{n(n^2-1)}\sum_{k=1}^{n-1}k(n^2-k^2)r_k}. \qquad (41)$$

Тоді незміщена оцінка дисперсії відхилень (34) набуває вигляду

$$S^2_{v,н.зм.}=\frac{S_v^2}{1-\frac{2}{n_{eff.1}}+\frac{1}{n_{eff.2}}}= \qquad (42)$$

$$=S_v^2\frac{n_{eff.1}}{n_{eff.1}-2+n_{eff.1}/n_{eff.2}}$$

Підставляючи у (23) замість s_v^2 її незміщену оцінку (42), отримуємо вираз для оцінки дисперсії коефіцієнта зміщення лінії

$$S^2_{b_0,kor}=\frac{S_v^2}{n}\frac{n_{eff.1}}{n_{eff.1}-2+n_{eff.1}/n_{eff.2}}\left(1+\frac{2}{n}\sum_{k=1}^{n-1}(n-k)r_k\right)=$$

$$=\frac{S_v^2}{n_{eff.1}-2+n_{eff.1}/n_{eff.2}}. \qquad (43)$$

Оскільки згідно з використаними позначеннями

$$\frac{1}{n}\left(1+\frac{2}{n}\sum_{k=1}^{n-1}r_k\sum_{j=1}^{n-k}\frac{x'_j x'_{k+j}}{S_x^2}\right)=\frac{1}{n}\left(1+\frac{2}{n}\sum_{k=1}^{n-1}(n-k)r_k-\right.$$

$$\left.-\frac{2}{n}\sum_{k=1}^{n-1}r_k\left((n-k)-\sum_{j=1}^{n-k}\frac{x'_j x'_{k+j}}{S_x^2}\right)\right)=\frac{1}{n_{eff.1}}-\frac{1}{n_{eff.2}}, \qquad (44)$$

то у разі корельованих результатів вимірювань оцінка дисперсії коефіцієнта нахилу лінії описується залежністю

$$S^2_{a_1,kor}=\frac{S_v^2}{n\cdot S_x^2}\cdot\frac{n_{eff.1}}{n_{eff.1}-2+n_{eff.1}/n_{eff.2}}\left(1+\frac{2}{n}\sum_{k=1}^{n-1}r_k\sum_{j=1}^{n-k}\frac{x'_j x'_{k+j}}{S_x^2}\right)=$$

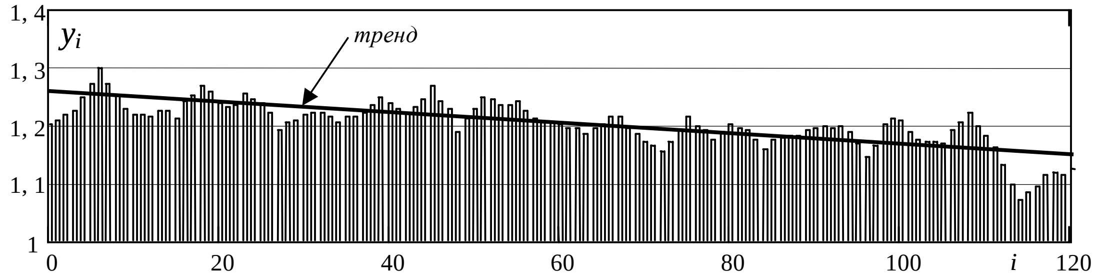
$$=\frac{S_v^2}{S_x^2}\frac{1-n_{eff.1}/n_{eff.2}}{n_{eff.1}-2+n_{eff.1}/n_{eff.2}}. \qquad (45)$$

Завдяки $\overline{x'}=0$ оцінки коефіцієнтів b_0 та a_1 взаємно некорельовані, тому, враховуючи (43) та (45), вираз для оцінки дисперсії прогнозованих значень вихідної величини набуває вигляду

$$s_y^2(x')=S^2_{b_0,kor}+S^2_{a_1,kor}x'^2=$$

$$=\frac{S_v^2}{n_{eff.1}-2+n_{eff.1}/n_{eff.2}}\left(1+\left(1-n_{eff.1}/n_{eff.2}\right)\cdot\frac{(x-\bar{x})^2}{S_x^2}\right). \qquad (46)$$

3. Приклад. Нехай зареєстровано $n = 121$ результатів спостережень напруги $y_i = u_i$, В, виміряних цифровим 4½ розрядним вольтметром у регулярних часових моментах:
1, 2200 1, 2080 1, 2186 1, 2263 1, 2497 1, 2725 1, 2981 1, 2731 1, 2500 1, 2286 1, 2181 1, 2183 1, 2162 1, 2247 1, 2253 1, 2108 1, 2409 1, 2529 1, 2696 1, 2577 1, 2397 1, 2300 1, 2341 1, 2562 1, 2449 1, 2378 1, 2203 1, 1920 1, 2056 1, 2092 1, 2198 1, 2227 1, 2210 1, 2134 1, 2064 1, 2138 1, 2154 1, 2220 1, 2352 1, 2479 1, 2385 1, 2277 1, 2206 1, 2320 1, 2466 1, 2679 1, 2412 1, 2279 1, 1897 1, 2123 1, 2291 1, 2498 1, 2450 1, 2343 1, 2356 1, 2420 1, 2239 1, 2101 1, 2057 1, 2044 1, 2011 1, 1940 1, 1941 1, 1836 1, 1956 1, 2002 1, 2159 1, 2142 1, 1963 1, 1840 1, 1726 1, 1657 1, 1553 1, 1726 1, 1932 1, 2146 1, 1983 1, 1904 1, 1736 1, 1874 1, 2003 1, 1950 1, 1911 1, 1754 1, 1594 1, 1748 1, 1799 1, 1817 1, 1816 1, 1907 1, 1937 1, 1982 1, 1956 1, 1977 1, 1868 1, 1684 1, 1455 1, 1648 1, 2019 1, 2126 1, 2086 1, 1885 1, 1760 1, 1729 1, 1706 1, 1692 1, 1921 1, 2036 1, 2229 1, 1996 1, 1810 1, 1609 1, 1314 1, 0975 1, 0704 1, 0845 1, 0954 1, 1146 1, 1172 1, 1148 1, 1263.



Зареєстровані результати спостережень напруги

У результатах спостережень необхідно виділити (щодо середини вибірки) лінійний тренд, знайти найкращу оцінку скоригованого результату – значення лінії для середини інтервалу спостережень, а також ураховуючи кореляцію між результатами спостережень $\left(r_k = e^{-0,455 \cdot k}\right)$, оцінити його стандартну непевність та стандартну непевність прогнозованих значень напруги.

Розв'язання:

1. На рисунку подано результати спостережень напруги, з якого видно наявність істотної часової зміни (лінійний тренд) напруги.

2. Приймаючи за значення аргументу відповідний номер результату, відповідно до (4) знаходимо:

$$S_x^2 = 1{,}220 \cdot 10^3, \; S_y^2 = 1{,}5495 \cdot 10^{-3} B^2,$$

$$r_{xy} = -0{,}7628 .$$

3. За цими значеннями знаходимо значення найкращого результату вимірювання для середини інтервалу спостережень та коефіцієнт нахилу (лінійного часового тренду) (3)

$$b_0 = 1{,}202867 \; B, \; a_1 = -8{,}596748 \cdot 10^{-4} \; 1/B .$$

4. На основі (36) і (41) та з урахуванням функції автокореляції визначаємо ефективні кількості результатів

$$n_{eff.1} \approx 29, \; n_{eff.2} \approx 843 .$$

5. За цими значеннями знаходимо незміщену оцінку дисперсії відхилень

$$S_{v,н.зм.}^2 \approx 4{,}57 \cdot 10^{-3} \; B^2 .$$

6. Далі визначаємо незміщені оцінки дисперсії результату вимірювання (значення коефіцієнта b_0) та коефіцієнт нахилу – лінійного тренду a_1:

$$S_{b_0,kor}^2 = 1{,}592 \cdot 10^{-4} \; B^2,$$

$$\text{та } S_{a_1,kor}^2 \approx 1{,}260 \cdot 10^{-7} \; 1/B^2$$

7. Якщо не враховувати взаємну кореляцію між результатами, то незміщена оцінка дисперсії результату вимірювання (значення коефіцієнта b_0) становила б відповідно

$$S_{b_0}^2 = 3{,}575 \cdot 10^{-5} \; B^2, \; S_{a_1}^2 \approx 2{,}930 \cdot 10^{-8} \; 1/B^2 .$$

Ці значення є більш ніж у 4 рази меншими за фактичні.

4. Висновки

1. Під час опрацювання результатів вимірювань методом лінійної регресії необхідно враховувати можливу взаємну кореляцію між почерговими випадковими результатами спостережень. Неурахування такої кореляції може призвести до хибних (занадто оптимістичних) оцінок шуканих величин.

2. Взаємна кореляція між почерговими випадковими результатами спостережень спричиняє зменшення ефективної кількості результатів спостережень, тим більше, чим тісніша кореляція, а це, своєю чергою, зумовлює погіршення оцінок дисперсії відхилень експериментальних точок від знайденої лінії, і, що найважливіше, оцінок дисперсії коефіцієнтів лінії і прогнозованих значень.

3. Для належного оцінювання впливу кореляції необхідне знання кореляційної функції, оцінку якої також можна знайти на підставі опрацювання результатів спостережень.

1. Крамер Г. Математические методы статистики. – М.: Мир, 1975. – 648 с. 2. Статистическая обработка результатов экспериментов на микро-ВМ и программируемых калькуляторах. / Ф.Ф. Костылев, П. В. Миляев, Ю. Д. Дорский и др. – Л.: Энергоатомиздат, 1991. – 304 с. 3. Львовский Е. Н. Статистические методы построения эмпирических формул. – М.: Высшая школа, 1988. – 239 с. 4. Дорожовець М., Стадник Б., Мотало В. та ін. Основи метрології і вимірювальна техніка: Підручник. – Т. 1. – Львів: Вид. Нац. ун-ту "Львівська політехніка", 2005. – 532 с. 5. Дорожовець М. Опрацювання результататів вимірювань. Навч. посібник. – Львів: Вид. НУ "Львівська політехніка", 2007. – 624 с. 6. Draper N.R. and Smith H. Applied Regression Analysis. Willey & Sons, New York, 1973. 7. Gillard J.W. An Historical Overview of Linear Regression with Errors in Both Variables // School of Mathematics, Senghenydd Road, Cardiff University, 2006. 8. Kendall M.G. and Stuart A. The Advanced Theory of Statistics // Volume Two. Charles Griffin and Co Ltd, London, Third edition, 1973. 9. Dorozhovets M. Niepewność liniowej regresji ortogonalnej // Pomiary, Automatyka, Kontrola. N9_bis, 2007, T.I, s.31-34. Kongres KKM-2007 w Krakowie. 10. Dorozhovets M. , Warsza Z. Propozycje rozszerzenia metod wyznaczania niepewności wyniku pomiarów wg Przewodnika GUM (1) // Uwzględnianie wpływu autokorelacji i nieadekwatności rozkładu wyników obserwacji w niepewności typu A. Pomiary, Automatyka, Robotyka. N1, 2007. – S. 16–25. 11. Dorozhovets M., Warsza Z. Methods of upgrading the uncertainty of type A evaluation (2). Elimination of the influence of autocorrelation of observations and choosing the adequate distribution // Proceedings of 15-th IMEKO TC4 Symposium. Novelties of Electrical Measurement and Instrumentation. Sep. 19-21, Jasi, Romania. 12. Dorozhovets M. Regresja liniowa uwzględniająca niepewności pomiaru obydwu wielkości // Materiały konferencji "Podstawowe problemy metrologii PPM-2008". Sucha Beskidzka, 12–15.05.2008. – S.25–28. 13. Guide of the Expression of Uncertainty in Measurement. ISO 1993, 1995. 14. Dorozhovets M. Wybrane problemy praktycznej oceny błędów oraz niepewności wyników pomiaru // Elektrotechnika, z. 29. Zeszyty Naukowe Politechniki Rzeszowskiej). 2006. – S. 9–44.