

**Фостенко К.В.**

*студент;*

**Ковтонюк І.Ю.**

*старший викладач кафедри вищої математики*

*Національний авіаційний університет*

## **ЗАСТОСУВАННЯ РЕГРЕСІЙНОГО АНАЛІЗУ ДЛЯ ЗНАХОДЖЕННЯ ТА УСУНЕННЯ ЗМІННИХ СИСТЕМАТИЧНИХ ПОХИБОК**

У даній статті розглянуто алгоритм та практичне застосування методу регресійного аналізу як способу перевірки результатів вимірювання на наявність систематичної зростаючої або спадної складової похибки, а також її усунення.

Досить часто в результаті вимірювання можна отримати значення, в яких присутня систематична похибка, зумовлена різними чинниками, наприклад, часом прогрівання засобу вимірювальної техніки (ЗВТ).

Залежно від характеру зміни значення в часі розрізняють сталі (систематичні) та змінні (прогресуючі, регулярні та випадкові) похибки.

Одним із методів виявлення змінної систематичної похибки є регресійний аналіз ( залежність однієї величини від іншої). Його використання проводиться на основі побудованого рівняння регресії і визначення внеску кожної незалежної змінної у варіацію досліджуваної (прогнозованої) величини.

Розглянемо алгоритм усунення змінної систематичної похибки: нехай у результаті вимірювання отримано  $N$  кількість значень величини  $Y_i$ , у моменти часу  $T_i$ , де  $i = 0, \dots, N$ .

- 1) Визначення наявності кореляції

$$r = \frac{\sum_{i=0}^{N-1} ((T_i - x_{mean}) \cdot (Y_i - y_{mean}))}{\sqrt{\sum_{i=0}^{N-1} (T_i - x_{mean})^2} \cdot \sqrt{\sum_{i=0}^{N-1} (Y_i - y_{mean})^2}}, \quad (1)$$

$$\text{де} \quad x_{mean} = \frac{1}{N} \sum_{i=0}^{N-1} T_i, \quad y_{mean} = \frac{1}{N} \sum_{i=0}^{N-1} Y_i \quad (2)$$

2) Перевірка значущості коефіцієнта кореляції

$$t = \left| \frac{r \cdot \sqrt{N-2}}{\sqrt{1-r^2}} \right| \quad (3)$$

Якщо  $t > t_{st}(P, N-2)$ , то кореляція значуща; при  $t \leq t_{st}(P, N-2)$ , то кореляція незначуща, де  $t_{st}(P, N-2)$  - коефіцієнт Стьюдента з  $N-2$  числом степенем вільності,  $P$  - довірна імовірність.  $P$  приймають за 0.95.

3) Оцінювання коефіцієнтів регресії

$$k^* = \frac{N \cdot \sum_{i=0}^{N-1} (T_i \cdot Y_i) - \sum_{i=0}^{N-1} T_i \cdot \sum_{i=0}^{N-1} Y_i}{N \cdot \sum_{i=0}^{N-1} T_i^2 - (\sum_{i=0}^{N-1} T_i)^2} \quad (4)$$

$$b^* = \frac{\sum_{i=0}^{N-1} Y_i - \sum_{i=0}^{N-1} (T_i \cdot k^*)}{N} \quad (5)$$

4) Оцінювання значущості коефіцієнтів регресії

$$S = \sqrt{\frac{1}{N-2} \cdot \sum_{i=0}^{N-1} (Y_i - b^* - k^* \cdot T_i)^2} \quad (6)$$

$$S_b = S \cdot \sqrt{\frac{1}{N} + \frac{x_{mean}^2}{\sum_{i=0}^{N-1} (T_i - x_{mean})^2}} \quad (7)$$

$$S_k = \frac{S}{\sqrt{\sum_{i=0}^{N-1} (T_i - x_{mean})^2}} \quad (8)$$

$$t_b = \left| \frac{b^*}{S_b} \right|, t_k = \left| \frac{k^*}{S_k} \right| \quad (9)$$

Якщо  $t_b > t_{st}(P, N-2)$ , то коефіцієнт  $b$  значущий; при  $t_k > t_{st}(P, N-2)$ , то коефіцієнт  $k$  значущий. При  $t_b \leq t_{st}(P, N-2)$ , коефіцієнт  $b$  незначущий, та при  $t_k \leq t_{st}(P, N-2)$  коефіцієнт  $k$  незначущий.

Рівняння регресії має вигляд:  $f^*(T) = k^* \cdot T_i + b^*$ .

5) Оцінювання адекватності моделі регресії

$$F = \frac{(N-2) \cdot \sum_{i=0}^{N-1} (y_{mean} - f^*(T_i))^2}{\sum_{i=0}^{N-1} (Y_i - f^*(T_i))^2} \quad (10)$$

Якщо  $F > F(P, 1, N-2)$ , то модель адекватна. При  $F \leq F(P, 1, N-2)$ , модель вважається неадекватною, де  $F(P, 1, N-2)$  - квантиль  $F$ -розподілу з довірчою імовірністю  $P = 0.975$  та ступенями свободи  $1$  та  $N-2$ .

б) Внесення поправки.

Обчислення оціненої поправки  $f^*_{correct}(T) = k \cdot T_i$ .

У результаті отримуємо виправлений результат вимірювання:

$$Y_{correct} = Y_i - f^*_{correct}(T)$$

Результати вимірювань величини  $Y_i$  до  $i$  після застосування алгоритму наведено в табл.1 і зображено на рис.1:

Таблиця 1

$Y_i$ випр	$Y_i$	$t_i$
2,848	2,901	0,00
2,856	2,957	0,05
2,857	3,016	0,10
2,860	3,069	0,15
2,860	3,145	0,20
2,860	3,203	0,25
2,868	3,254	0,30
2,872	3,307	0,35
2,875	3,416	0,40
2,878	3,505	0,45

продовження таблиці 1

$Y_i$ випр	$Y_i$	$t_i$
2,883	3,569	0,50
2,885	3,623	0,55
2,888	3,675	0,60
2,891	3,731	0,65
2,891	3,787	0,70
2,901	3,844	0,75
2,901	3,910	0,80
2,913	3,972	0,85
2,914	4,038	0,90
2,960	4,207	0,95

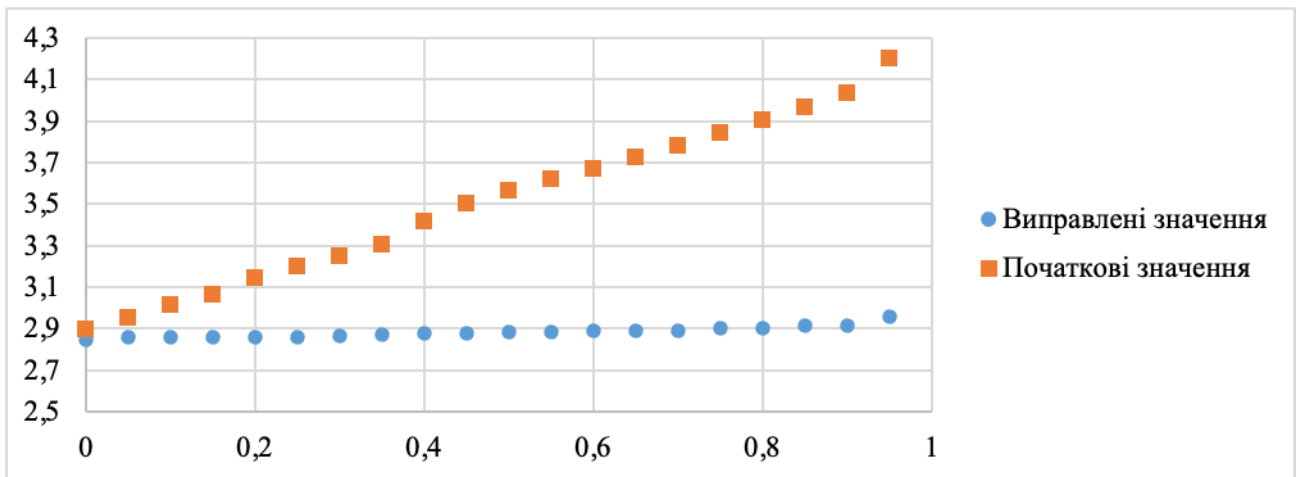


Рис. 1. Порівняння виправлених та початкових значень

Застосування регресійного аналізу при вимірюванні дало змогу майже усунути систематичну складову похибки. Але також необхідно враховувати, що внесення поправки збільшує випадкову складову похибки. Тому потрібно вирішити, чи вагомий внесок робить її систематична складова.

#### Список використаних джерел:

1. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. Множественная регрессия. — 3-е изд. — М.: «Диалектика», 2007.