

ПОРІВНЯННЯ РІЗНИХ СПОСОБІВ ВИРІВНЮВАННЯ СТАТИСТИЧНИХ ДАНИХ МЕТОДОМ НАЙМЕНШИХ КВАДРАТІВ

Науковий керівник: к.ф.-м.н., доцент Демчишин О.І.

Якщо вибірка змінних y та x задана статистичними даними, то конфігурація точок $(x_i; y_i)$, де $i \in \mathbf{N}_n$, в прямокутній системі координат дозволяє висунути гіпотезу про вид функції, а метод найменших квадратів дає можливість реалізувати вибрану функціональну залежність.

За цим методом найбільш вживаними є вирівнювання статистичних даних за допомогою: прямої лінії – $\hat{y}_1 = a_1 + b_1x$; параболи – $\hat{y}_2 = a_2 + b_2x + c_2x^2$;

гіперболи – $\hat{y}_3 = a_3 + \frac{b_3}{x}$; показникової кривої – $\hat{y}_4 = a_4 \cdot b_4^x$.

Для оцінки відхилення отриманих значень $\hat{y}_{\omega i}$ ($\omega = 1; 2; 3; 4$) із функціональної залежності і статистичних даних y_i ($i \in \mathbf{N}_n$) пропонується використати суму квадратів цих відхилень і проводити оцінювання за допомогою параметрів:

$$\sigma_1 = \frac{\sqrt{\sum_{i=1}^n (\hat{y}_{1i} - y_i)^2}}{n}; \quad \sigma_2 = \frac{\sqrt{\sum_{i=1}^n (\hat{y}_{2i} - y_i)^2}}{n}; \quad \sigma_3 = \frac{\sqrt{\sum_{i=1}^n (\hat{y}_{3i} - y_i)^2}}{n}; \quad \sigma_4 = \frac{\sqrt{\sum_{i=1}^n (\hat{y}_{4i} - y_i)^2}}{n}.$$

Для прикладу, розглянемо сукупність спостережень

x	1	2	3	4	5	6	7	8	9
y	3,2	1,8	1,1	0,3	-0,2	-2,2	-2,3	-4,1	-5

Вирівнювання за допомогою прямої лінії $\hat{y}_1 = a_1 + b_1x$ і гіперболи $\hat{y}_2 = a_2 + \frac{b_2}{x}$ із систем рівнянь

$$\left\{ \begin{array}{l} na_1 + b_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i, \\ a_1 \sum_{i=1}^n x_i + b_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i. \end{array} \right. \quad \text{та} \quad \left\{ \begin{array}{l} na_2 + b_1 \sum_{i=1}^n z_i = \sum_{i=1}^n y_i, \\ a_2 \sum_{i=1}^n x_i + b_2 \sum_{i=1}^n z_i^2 = \sum_{i=1}^n z_i y_i, \end{array} \right. \quad \left(z_i = \frac{1}{x_i} \right)$$

дає нам рівняння: $\hat{y}_1 = 4.16 - x$; $\hat{y}_2 = -23 + \frac{1}{x}$.

Оцінювання відхилень за запропонованим способом показує, що $\sigma_1 = 0.115$ і $\sigma_2 = 7.33$.

Таким чином, робимо висновок: для заданих статистичних даних оптимальнішим є функціональна залежність, отримана за допомогою вирівнювання прямою лінією.

1. Кальницький Л. А. и др. Специальный курс высшей математики для вузов. Учеб. пособие. М., «Высшая школа», 1976. 388 с.