

## РОЗРОБКА КОРИСНИХ КОПАЛИН

УДК 331.45

О.Є. Кружилко, к.т.н., доц.

В.В. Майстренко, с.н.с.

Ю.О. Полукаров, м.н.с.

Національний науково-дослідний інститут охорони праці

## МЕТОДИКА ЗАСТОСУВАННЯ РЕГРЕСІЙНОГО АНАЛІЗУ ДЛЯ МОДЕЛЮВАННЯ ПОКАЗНИКІВ УМОВ ПРАЦІ ЗВАРЮВАЛЬНОГО ВИРОБНИЦТВА

У статті запропонована методика застосування регресійного аналізу для моделювання показників умов праці на прикладі зварювального виробництва. Наведені основні етапи побудови математичної моделі та результати обчислень розробленого програмного комплексу.

Для аналізу і прогнозування стану охорони праці використовуються методи математичного моделювання. На основі зібраних статистичних даних розраховується математична модель, яка встановлює функціональну залежність показника стану охорони праці від множини чинників, що на них впливають. Як показав проведений аналіз, метод математичного моделювання в сфері охорони праці знайшов застосування здебільшого для прогнозування стану виробничого травматизму та його наслідків [1]–[3]. З використанням результатів моделювання отримано оцінку впливу чинників на показники травматизму та одержано рекомендації щодо планування працезахоронних заходів.

Зважаючи на позитивний досвід застосування методів математичного моделювання для вирішення широкого класу задач управління охороною праці, слід зазначити деякі аспекти, що потребують вирішення. Вибір методу моделювання, і, як наслідок, вигляд математичної моделі залежить не тільки від природи об'єкта моделювання, але і від тих задач, для вирішення яких вона створюється. Тому перш ніж приступити до моделювання, необхідно визначити коло задач, які планується вирішувати.

Як свідчить статистика останніх років, спостерігається тенденція зростання рівня професійної захворюваності працівників різних професій. Оскільки задача моделювання та прогнозування рівня професійної захворюваності (на відміну від рівня виробничого травматизму) в наукових працях та публікаціях не знайшла широкого висвітлення, необхідно розробити (адаптувати) алгоритм моделювання саме цього показника з урахуванням особливостей множини чинників, що впливають на це. Зважаючи на те, що при моделюванні необхідно виконувати велику кількість математичних операцій, необхідно розробити програмний комплекс для автоматизації розрахунків та зберігання результатів моделювання.

Розглянемо методику застосування регресійного аналізу для вирішення поставленої задачі.

В загальному випадку математична модель, що визначає залежність показника від множини чинників, має вигляд:

$$Y = Y(X_1, \dots, X_n), \quad (1)$$

де  $Y$  – модельований показник (параметр);  $X_1, \dots, X_n$  – чинники, що впливають на модельований показник;  $n$  – число чинників.

Статистичні дані про чинники  $X_1, \dots, X_n$  заносяться до таблиці.

Таблиця

Масив статистичних даних для побудови моделі

Чинники, що характеризують умови праці			Показник
$X_1$	...	$X_n$	$Y$
$x_{11}$	...	$x_{n1}$	$y_1$
...	...	...	...
$x_{1m}$	...	$x_{nm}$	$y_m$

У наведеній таблиці маємо:  $x_{ij}$  – значення  $j$ -го чинника в  $i$ -й момент спостереження,  $R_{X,Y}$ ,  $i = 1..m$ ;  $y_j$  – значення показника в  $j$ -й момент спостереження;  $m$  – число статистичних спостережень.

Як показник для умов зварювального виробництва найбільш важливим є показник професійної захворюваності робітників. Як чинники використовуються технологічні, організаційні, економічні, санітарно-гігієнічні, медико-соціальні та інші.

Для застосування множинного регресійного аналізу повинні виконуватись такі умови: при кожному наборі значень  $X_i$  величина  $Y$  розподілена нормально та дисперсія величини  $Y$  є постійною величиною. При цьому шукана модель має вигляд:

$$Y = a_0 + \sum_{i=1}^n a_i X_i + \sum_{i=1}^n \sum_{j=1}^n a_{ij} X_i X_j + \dots \epsilon. \quad (2)$$

Підготовка даних до моделювання є першим етапом дослідження впливу чинників, що характеризують умови праці, на показники. Перш ніж приступити до розрахунку коефіцієнтів рівнянь регресії, необхідно виключити з розгляду ті чинники, що не мають істотного впливу на модельований показник. Для цього необхідно розрахувати коефіцієнти парної кореляції  $R_{X_i Y}$  між показником і чинниками та перевірити їх значимість, тобто встановити міру впливу кожного окремого чинника на показник.

Для кожного чинника  $X_i$ , а також для показника  $Y$  обчислюємо їх середні значення:

$$\bar{X}_i = \frac{1}{m} \sum_j X_{ij}, \quad \bar{Y} = \frac{1}{m} \sum_j Y_j. \quad (3)$$

Крім того, обчислюємо такі значення.

Суму квадратів відхилень:

$$S_{X_i X_i} = \sum_j (X_{ij} - \bar{X}_i)^2, \quad S_{YY} = \sum_j (Y_j - \bar{Y})^2. \quad (4)$$

Стандартне відхилення:

$$S_{X_i} = \sqrt{S_{X_i X_i} / (m-1)}, \quad S_Y = \sqrt{S_{YY} / (m-1)}. \quad (5)$$

Сума добутків відхилень:

$$S_{X_i Y} = \sum_j (X_{ij} - \bar{X}_i)(Y_j - \bar{Y}), \quad S_{X_i X_j} = \sum_k (X_{ik} - \bar{X}_i)(X_{jk} - \bar{X}_j). \quad (6)$$

Коефіцієнти кореляції:

$$R_{X_i Y} = \frac{S_{X_i Y}}{\sqrt{S_{X_i X_i} S_{YY}}}, \quad R_{X_i X_j} = \frac{S_{X_i X_j}}{\sqrt{S_{X_i X_i} S_{X_j X_j}}}. \quad (7)$$

Після перевірки значимості коефіцієнтів кореляції  $R_{X_i Y}$  з рівняння регресії виключаються чинники  $X_i$ , для яких  $R_{X_i Y} < \Delta$ , де  $\Delta$  – поріг значимості коефіцієнтів кореляції. Необхідно зазначити, що у випадку недостатньої кількості точок (число точок повинне бути більше числа чинників) необхідно ранжувати  $X_i$  за убаванням коефіцієнта кореляції  $R_{X_i Y}$  і відібрати чинники для побудови моделі. Далі оцінюємо кореляцію між  $X_i$  і  $X_j$ . Якщо вони лінійно зв'язані між собою ( $R_{X_i X_j} \Rightarrow 1$ ), то в цьому випадку один з чинників нічого істотного в порівнянні з іншими не вносить у рівняння, отже слід включати з один з таких чинників з множини початкових даних.

При практичній роботі виправдовує себе процедура відсівання. Відповідно до цієї процедури оцінка початкових даних ґрунтується на аналізі коефіцієнтів парної кореляції  $R_{X_i Y}$ . За результатами розрахунків множина чинників  $X$  слід сортувати в порядку убавання значення коефіцієнтів парної кореляції. Графічне зображення цієї процедури на прикладі масиву з 10 чинників представлено на рис. 1. Якщо число чинників перевищує число точок спостереження ( $n > m$ ), максимально припустимим числом чинників для побудови математичних моделей буде  $m - 1$  [4]. В цьому випадку з відсортованого масиву даних відбираються перші за убаванням коефіцієнта кореляції  $R_{X_i Y}$   $m - 1$  чинників.

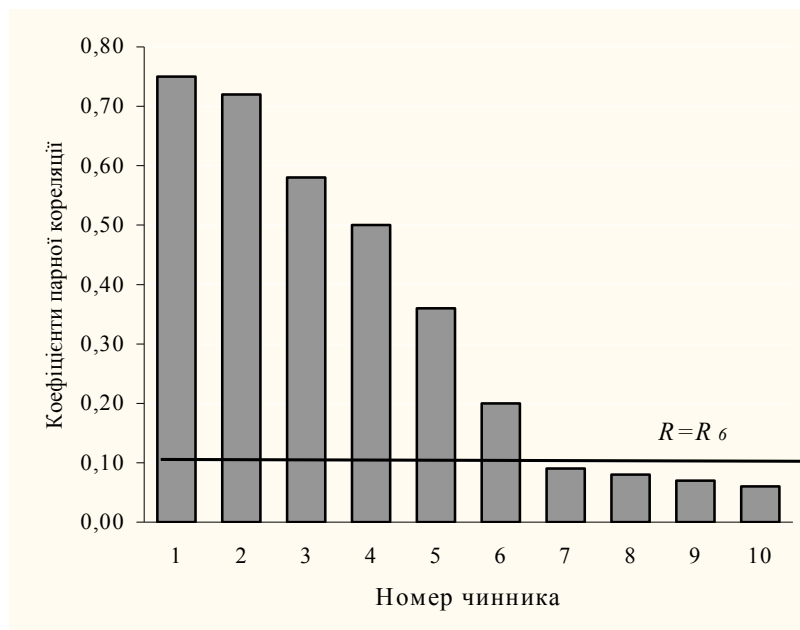


Рис. 1. Графічне зображення процедури відсіювання чинників за коефіцієнтом кореляції

У випадку, якщо число чинників менше або дорівнює числу точок спостереження ( $n \leq m$ ), процедура оцінки чинників здійснюється в процесі ітераційної побудови моделей, починаючи з одного чинника, що має найбільший коефіцієнт кореляції  $R_{X_1Y}$ . На кожному наступному кроці в модель додається по одному новому чиннику в порядку за убаванням значень коефіцієнтів кореляції  $R_{X_iY}$ . На кожному кроці точність розрахованої моделі оцінюється за коефіцієнтом множинної детермінації (20). Якщо розрахована на  $i$ -му кроці точність моделі не перевищує точність на кроці  $i - 1$ , процес побудови моделей закінчується, а чинники, які не увійшли в модель можна вважати такими, що не мають істотного впливу на модельований показник. На практиці точність моделі вважається задовільною, якщо коефіцієнт множинної детермінації не менше 0,95 [5].

При використанні лінійного багатомірного регресійного аналізу ставиться задача одержання оптимальної оцінки параметра  $Y$ , виходячи з набору чинників  $X_i$ . Розглянемо задачу визначення рівняння лінійної регресії.

Припустимо, що в результаті обробки статистичних даних отримана залежність, яка має вигляд поліному першого степеня:

$$Y = a_0 + \sum_{i=1}^n a_i X_i . \tag{8}$$

Задача полягає в перевірці правомірності припущення про наявність залежності виду (2) і перебуванні коефіцієнтів рівняння (3.2)  $a_0, a_1, \dots, a_n$ , для яких досягається

$$\min_{a_0 \dots a_n} \left[ Y - \left( a_0 + \sum_{i=1}^n a_i X_i \right) \right]^2 . \tag{9}$$

Для спрощення розрахунків коефіцієнтів регресії необхідно нормувати значення  $X_i$  і  $Y$  та перейти до нових змінних:

$$W_{ij} = \frac{X_{ij} - \bar{X}_i}{S_{X_i}}, U_j = \frac{Y_j - \bar{Y}}{S_Y} . \tag{10}$$

Слід зазначити, що перехід до нових змінних здійснюється з метою одержання стандартних середніх значення змінних  $\bar{W}_i = 0$  і  $U = 0$ , а дисперсія  $S_{W_i} = S_U = 1$ .

Після переходу до нових змінних одержимо лінійне рівняння регресії:

$$U = \beta_1 W_1 + \beta_2 W_2 + \dots + \beta_n W_n , \tag{11}$$

де  $\beta_i$  – нові коефіцієнти регресії, лінійно зв'язані з коефіцієнтами  $a_i$  наступними співвідношеннями:

$$a_0 = \bar{Y} - \sum_{i=1}^n a_i X_i, \quad a_i = \beta_i \frac{S_Y}{S_{X_i}}.$$

Причому  $R_{X_i X_j} = R_{W_j W_i}$ ;  $R_{Y X_i} = R_{U W_i}$ .

Коефіцієнти  $\beta_i$  визначаються із системи рівнянь:

$$\begin{aligned} R_{Y X_1} &= \beta_1 + \beta_2 R_{X_2 X_1} + \beta_3 R_{X_3 X_1} + \dots + \beta_n R_{X_n X_1} \\ R_{Y X_2} &= \beta_1 R_{X_1 X_2} + \beta_2 + \beta_3 R_{X_3 X_2} + \dots + \beta_n R_{X_n X_2}, \\ &\dots \\ R_{Y X_m} &= \beta_1 R_{X_1 X_m} + \beta_2 R_{X_2 X_m} + \beta_3 R_{X_3 X_m} + \dots + \beta_n \end{aligned} \quad (12)$$

де  $R_{XY}$  – вектор коефіцієнтів кореляції між  $X_i$  та  $Y$ ;

$R_{XX}$  – матриця кореляції між чинниками;

$B$  – вектор коефіцієнтів  $\beta_i$ .

Для розв'язку системи рівнянь (13) необхідно знайти зворотну матрицю  $R_{XX}^{-1}$ :

$$R_{XX}^{-1} R_{YX} = R_{XX}^{-1} R_{XX} B. \quad (13)$$

Розв'язок має вигляд:

$$B = R_{XX}^{-1} R_{YX}, \quad \beta_i = \sum_{j=1}^n R_{XX_{ij}}^{-1} R_{YX_j}. \quad (14)$$

Для перебування зворотної матриці можна використовувати наступну процедуру.

Будується допоміжна матриця  $C = (c_{ik})$ ,  $i \leq k$ , що є наддіагональною трикутною матрицею [5] і елементи якої обчислюються за формулами:

$$\begin{aligned} c_{ii} &= \sqrt{R_{X_i X_i} - c_{1i}^2 - c_{2i}^2 - \dots - c_{i-1,i}^2}, \\ c_{ik} &= (R_{Z_j Z_k} - c_{1i} c_{1k} - c_{2i} c_{2k} - \dots - c_{i-1,i} c_{i-1,k}) c_{ii} \quad \text{для } i < k. \end{aligned} \quad (15)$$

Побудова матриці  $C$  виконується в такий спосіб. Діагональні елементи  $c_{ii}$  визначаються в результаті витягу квадратного кореня з різниці  $R_{X_j X_j}$  і сум квадратів елементів  $c_{ki}$ , отриманих раніше. Елементи  $c_{ki}$ , що віднімаються з  $R_{X_k X_j}$ , розташовані в стовпцях матриці над діагональними елементами, що потрібно одержати. Інші елементи рядка визначаються, віднімаючи із  $R_{X_k X_j}$  скалярний добуток двох векторів-стовпців матриці  $C$ , що відповідають індексам  $i$  і  $k$ , і результат ділиться на  $c_{ii}$ . При утворенні скалярного добутку не враховується елемент, що знаходиться на цьому новому рядку. Для наддіагональної матриці  $C$  має місце рівність  $R = C'C$ . Таким чином, матрицю  $R$  можна подати у вигляді добутку двох трикутних матриць, транспонованих відносно одна до одної. Розкладання на такі складові можна робити з усіма кореляційними матрицями. Процедура обчислень від цього не змінюється, тому що уявні елементи в остаточних результатах зникають. Аналогічно обчислюємо значення  $c_i$  за формулою:

$$c_i = \frac{R_{X_j X_i} - \sum_{j=1}^{i-1} c_{ji} c_j}{c_{ii}}. \quad (16)$$

Визначення невідомих  $\beta_1, \beta_2 \dots \beta_n$  виконується за матрицею  $C$  в зворотній послідовності:

$$\begin{aligned} \beta_n &= c_n c_{nn}; \\ \beta_{n-1} &= (c_{n-1} - c_{n-1n} \beta_n) c_{n-1n-1}; \\ \beta_1 &= (c_1 - c_{1n} \beta_n - c_{1n-1} \beta_{n-1} - c_{12} \beta_2) c_{11}. \end{aligned} \quad (17)$$

Оцінка отриманої моделі розраховується з використанням  $\beta_i$  та  $R_{X_i Y}$ .

Для того, щоб установити правомірність припущення про лінійну залежність модельованої величини  $Y$  від чинників  $X_1, X_2, \dots, X_n$  обчислюється коефіцієнт множинної кореляції:

$$R = \sqrt{\beta_1 R_{YX_1} + \beta_2 R_{YX_2} + \dots + \beta_n R_{YX_n}}. \quad (18)$$

У залежності від кількості чинників і числа точок, за допомогою критеріїв Фішера і Стьюдента, встановлюється значимість коефіцієнта множинної кореляції  $R$  і коефіцієнтів регресії  $\beta_i$ .

Якщо  $R$  значно відрізняється від 0, то лінійний зв'язок існує, і припущення про наявність зв'язку між  $X_i$  і  $Y$  є вірним. У цьому випадку оцінюються коефіцієнти регресії  $\beta_i$ . Для їхньої оцінки визначається коефіцієнт множинної детермінації, що дорівнює квадрату коефіцієнта множинної кореляції:

$$R^2 = \sum_{i=1}^m \beta_i R_{YX_i} . \quad (19)$$

За коефіцієнтом множинної детермінації можна визначити, яка частина дисперсії величини  $Y$  зумовлена чинником  $X_i$ , оскільки внесок окремого чинника в коефіцієнт множинної кореляції визначається відповідними доданками. Не знижуючи точності моделі, можна виключити при оцінці якості моделі ті чинники, внесок яких буде незначним.

Оцінка якості побудованих моделей здійснюється з використанням спеціальних критеріїв (коефіцієнтів).

Коефіцієнт множинної кореляції:

$$R = \sqrt{R^2} = \sqrt{\sum_{i=1}^m \beta_i R_{YX_i}} . \quad (20)$$

За таблицею, наведеною в [6], перевіряємо значимість коефіцієнта множинної кореляції з урахуванням числа незалежних змінних.

Якщо виконується умова  $R > R_{таб}$ , то підтверджується гіпотеза про адекватність отриманої моделі.

Коефіцієнт  $1 - R^2$  визначає точність опису показника регресійною моделлю. Якщо  $1 - R^2 < \Delta$ , де  $\Delta$  – похибка обчислень, то модель використовується для подальшого аналізу. Доданки  $\beta_i R_{YX_i}$  – визначають внесок чинника  $X_i$  в параметр  $Y$ . Якщо  $|\beta_i R_{YX_i}| < \Delta$ , то чинник  $X_i$  виключається з подальшого аналізу.

Критерій Фішера ( $F$ -критерій).

Для використання  $F$ -критерію необхідно обчислити:

$$Y_{P_j} = a_0 + \sum_{i=1}^n a_i X_{ij}, \quad \bar{Y}_{P_j} = \frac{1}{n} \sum_{j=1}^N Y_{P_j}, \quad S_{Y_P Y_P} = \sum_{j=1}^N (Y_{P_j} - \bar{Y}_P)^2 . \quad (21)$$

Якщо  $S_{YY} > S_{Y_P Y_P}$  (5), то

$$F = \frac{S_{YY}}{S_{Y_P Y_P}} . \quad (22)$$

В іншому випадку:

$$F = \frac{S_{Y_P Y_P}}{S_{YY}} . \quad (23)$$

За таблицями [6] для заданого значення  $m$  знаходимо допустиме значення  $F_{дон}$ . Адекватність моделі вважається доведеною, якщо  $F \leq F_{дон}$ .

$t$ -критерій.

Для визначення значення  $t$ -критерію обчислюємо коефіцієнт кореляції:

$$R_{YY_P} = \frac{S_{YY_P}}{\sqrt{S_{YY} \cdot S_{Y_P Y_P}}} , \quad (24)$$

де  $S_{YY_P} = \sum_{j=1}^n (Y_j - \bar{Y})(Y_{P_j} - \bar{Y}_P)$ .

$t$ -критерій у цьому випадку дорівнює:

$$t = \frac{R_{YY_P} \sqrt{n-2}}{\sqrt{1-R_{YY_P}^2}} . \quad (25)$$

За таблицями [6] для заданого значення  $m$  знаходимо допустиме значення  $t_{дон}$ . Якщо  $t \geq t_{дон}$ , то адекватність моделі вважається доведеною.

Критерій Стьюдента дорівнює:

$$t = \frac{|\bar{Y}_P - \bar{Y}|}{\frac{S}{\sqrt{N}}} = \frac{|\bar{Y}_P - \bar{Y}|}{\sqrt{\sum_{j=1}^N (Y_{P_j} - \bar{Y}_P)^2}} \cdot \sqrt{\frac{N}{N-1}} , \quad (26)$$

$$\text{де } S^2 = \frac{1}{N-1} \sum_{j=1}^N (Y_j - \bar{Y}_P)^2 .$$

Якщо  $t \geq t_{\text{доп}}$  (де  $t_{\text{доп}}$  – допустиме значення критерію Стьюдента), то модель вважається адекватною.

У більшості випадків лінійні багаточинникові моделі не дозволяють з необхідним ступенем точності адекватно описувати аналізовані параметри. Тому приходиться вдаватись до побудови нелінійних моделей, причому іноді, якщо це можливо, задача зводиться до лінійного багаточинникового регресійного аналізу. У розглянуту модель вводяться нові чинники:

$$X_l = f(X_1, \dots, X_n), \quad l > n, \tag{27}$$

де функції  $f(X_1, \dots, X_n)$  мають вигляд:

$$X_i \cdot X_j ; X_i \cdot X_j \cdot X_k ; \ln X_i \dots \tag{28}$$

Як правило, при моделюванні параметрів, що характеризують стан охорони праці, досить використовувати нові чинники, що є добутком початкових чинників. Враховуючи необхідність виконання значної кількості математичних розрахунків при побудові моделей, розроблено програмний комплекс, який дозволяє автоматизувати виконання розрахунків. Розглянемо його основні можливості. Введення даних до програми здійснюється після натискання кнопки „дані”. Розрахунок коефіцієнтів парної кореляції (значення коефіцієнтів наведено в шапці таблиці поруч з відповідним чинником  $X_i$ ) та сортування масиву введених даних після натискання кнопки „сортувати стовпці” (рис. 2).

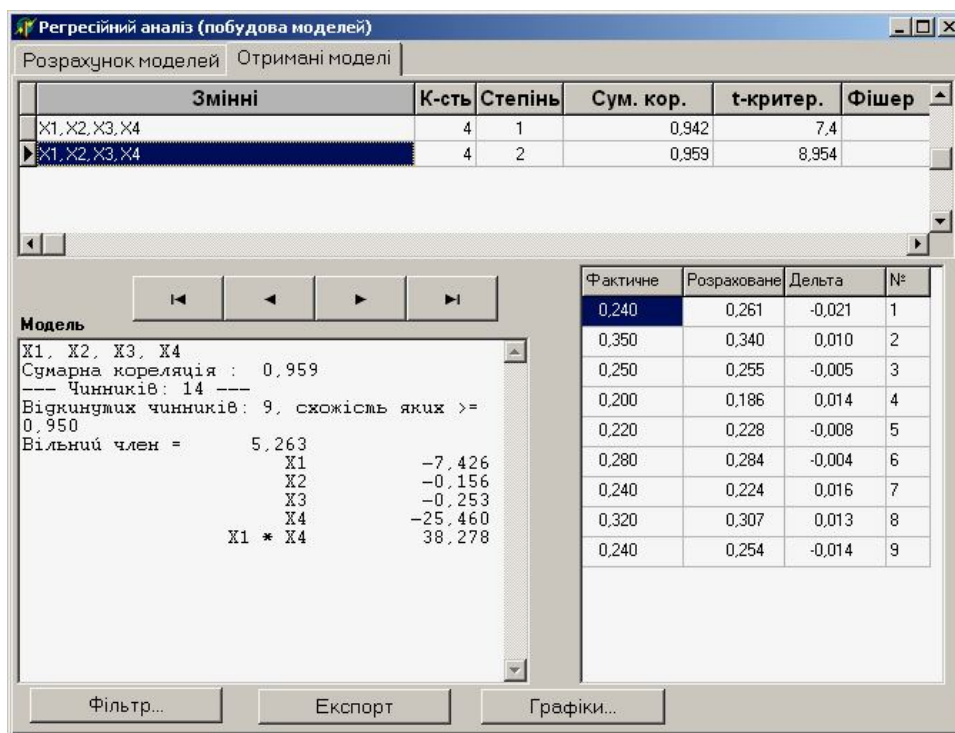


Рис. 2. Масив статистичних даних, підготовлений до моделювання

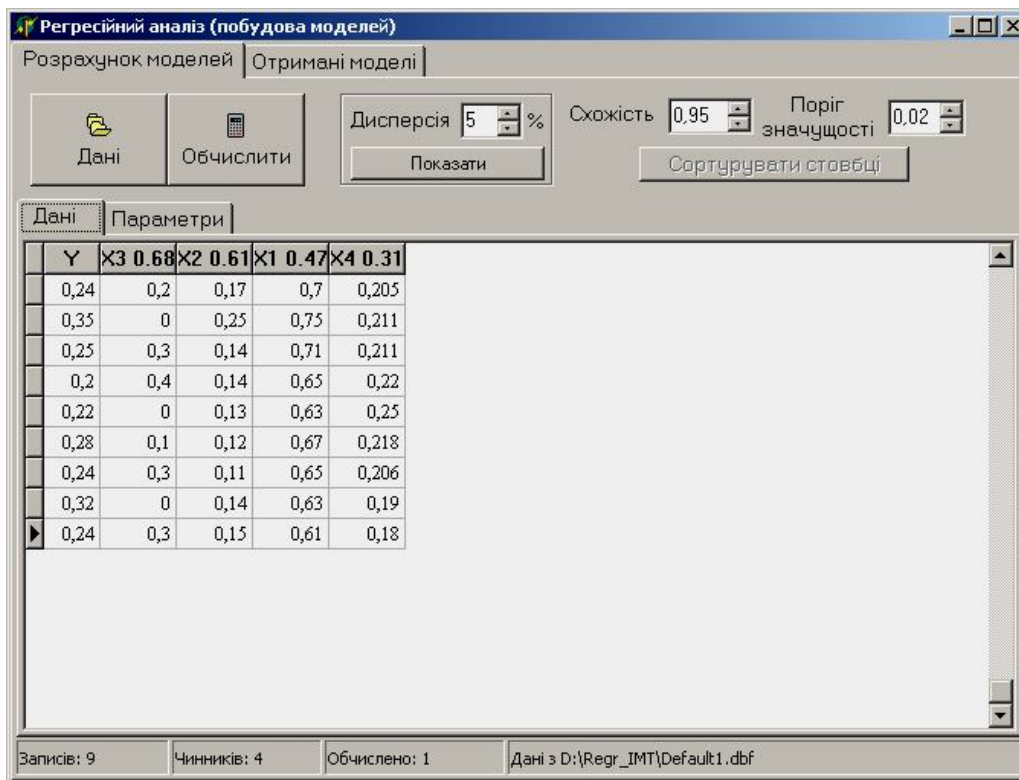


Рис. 3. Результати моделювання

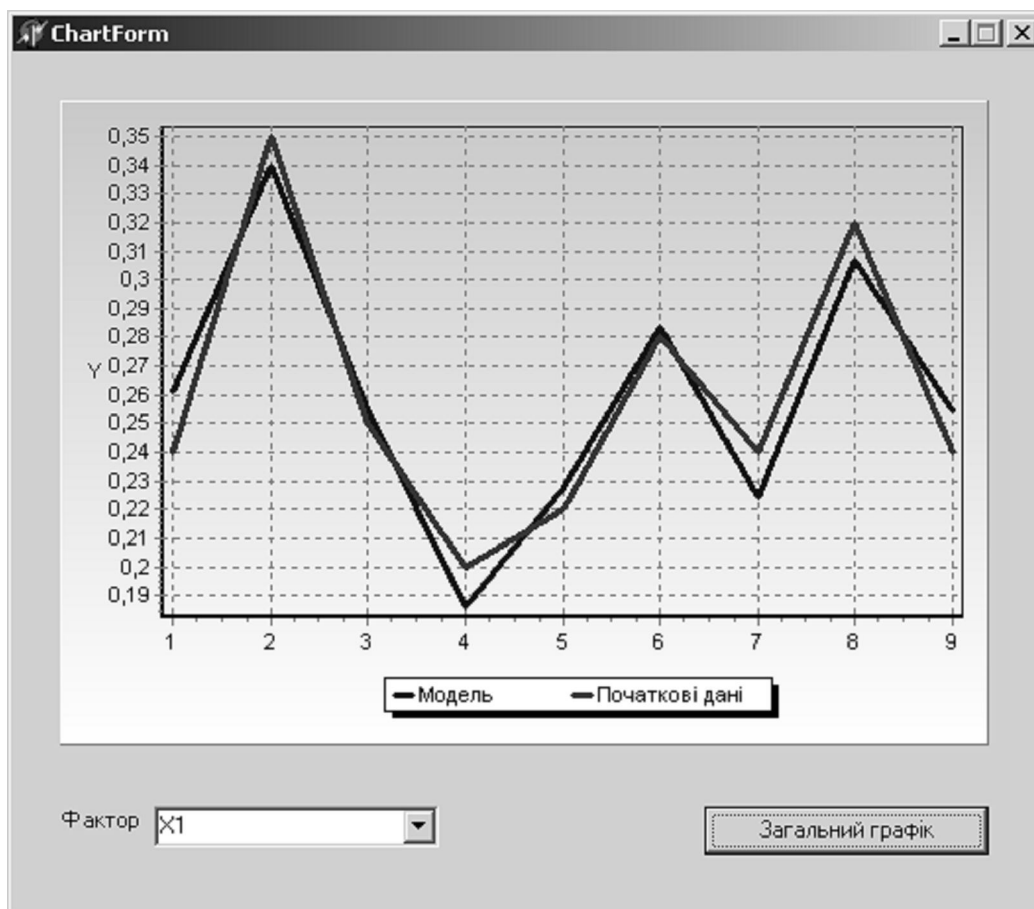


Рис. 4. Результати моделювання

Отже використання апарату математичного моделювання дозволить прогнозувати стан охорони праці при переході до виконання нових виробничих програм, зміни технологічних процесів, устаткування та умов виробництва.

#### ЛІТЕРАТУРА:

1. *Ткачук С.П., Ткачук К.Н., Кружилко О.Е., Майстренко В.В.* Моделирование процессов управления в области охраны труда / Разработка рудных месторождений. – Кривой Рог: КТУ. – 1997. – Вып. 61. – С. 113–119.
2. *Ступницька Н.В.* Математична модель стану виробничого травматизму в механоскладальному цеху. Оптимізація виробничих процесів і технічний контроль в машинобудуванні і приладобудуванні // Львів: Вісник Державного університету "Львівська Політехніка". – 1998. – № 321. – С. 103–105.
3. *Ткачук С.П.* Использование аппарата математического моделирования в системе управления охраной труда // Охрана труда. – 1999. – № 5. – С. 37–38.
4. *Ферстер Е., Ренц Б.* Методы корреляционного и регрессионного анализа. – М.: Финансы и статистика, 1983. – 302 с.
5. *Снапелев Ю.М., Старосельский В.А.* Моделирование и управление в сложных системах. – М.: Сов. радио, 1974. – 264 с.
6. *Митропольский А.К.* Техника статистических вычислений. – М.: Наука, 1971. – 576 с.

**КРУЖИЛКО Олег Євгенович** – кандидат технічних наук, доцент, старший науковий співробітник Національного науково-дослідного інституту охорони праці.

Наукові інтереси:

- експертні системи;
- професійні захворювання та травматизм.

**МАЙСТРЕНКО Володимир Володимирович** – старший науковий співробітник Національного науково-дослідного інституту охорони праці.

Наукові інтереси:

- аналіз виробничого травматизму;
- методи прогнозування травматизму.

**ПОЛУКАРОВ Юрій Олексійович** – молодший науковий співробітник Національного науково-дослідного інституту охорони праці.

Наукові інтереси:

- зварювальне виробництво;
- професійні захворювання зварювальників;
- засоби захисту зварювальників.

Подано 10.11.2004



**О.Є. Кружилко, В.В. Майстренко, Ю.О. Полукаров.** Методика застосування регресійного аналізу для моделювання показників умов праці зварювального виробництва

**О.Е. Кружилко, В.В. Майстренко, Ю.А. Полукаров.** Методика применения регрессионного анализа для моделирования показателей условий труда сварочного производства

**O.E. Kruzhilko, V.V. Mijstrenko, U.A. Polukarpov** Technique of application of the regressive analysis for modeling parameters of working conditions of welding manufacture //

УДК 331.45

**Методика застосування регресійного аналізу для моделювання показників умов праці зварювального виробництва // Кружилко О.Є., Майстренко В.В., Полукаров Ю.О. /**

У статті запропонована методика застосування регресійного аналізу для моделювання показників умов праці на прикладі зварювального виробництва. Наведені основні етапи побудови математичної моделі та результати обчислень розробленого програмного комплексу

УДК 331.45

**Методика применения регрессионного анализа для моделирования показателей условий труда сварочного производства // Кружилко О.Е., Майстренко В.В., Полукаров Ю.А.**

В статье предложена методика применения регрессионного анализа для моделирования показателей условий труда на примере сварочного производства. Приведены основные этапы построения математической модели и результаты вычислений разработанного программного комплекса.

УДК 331.45

**Technique of application of the regressive analysis for modeling parameters of working conditions of welding manufacture // Kruzhilko O.E., Mijstrenko V.V., Polukarpov U.A.**

In article is offered the strategy of using regressive analysis for modeling of factors of conditions of labour on example of welding production. They are brought main stages of building of mathematical model and results of calculations developed program complex.