МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ДНІПРОВСЬКИЙ ДЕРЖАВНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ

РЕГРЕСІЙНА МОДЕЛЬ З ОДНІЄЮ ВХІДНОЮ ЗМІННОЮ ДЛЯ СТАТИСТИЧНОЇ ОБРОБКИ ДАНИХ НА ПЕОМ

МЕТОДИЧНІ ВКАЗІВКИ

до практичної роботи з дисципліни
«Моделювання технологічних та фізичних процесів»
освітньо-наукової програми вищої освіти
(підготовка докторів філософії)
зі спеціальності 136 "Металургія"

Затверджено редакційно-видавничою секцією науково-методичної ради ДДТУ, від ___________ 2017 р., протокол № ___

Дніпродзержинськ
2017

Регресійна модель з однією вхідною змінною для статистичної обробки даних на ПЕОМ. Методичні вказівки до практичної роботи з дисципліни «Моделювання технологічних та фізичних процесів» освітньо-наукової програми вищої освіти (підготовка докторів філософії) зі спеціальності 136 “Металургія” / Пантейков С.П.- Кам’янське, ДДТУ, 2017.- 23 с.

Укладач: канд. техн. наук, доцент Пантейков С.П.

Відповідальний за випуск: канд. техн. наук, доцент Чубін К.І.

Рецензент: канд. техн. наук, доцент Самохвал В.М.

Затверджено на засідані кафедри металургії чорних металів
Протокол № 11 від 05 червня 2017 р.

Розглянуто процес розробки регресійної моделі з однією вхідною змінною для статистичної обробки результатів досліджень, наведений приклад. Викладені методичні вказівки щодо проведення розрахунків і оформлення завдання, наведені вхідні дані до розрахунків.

З М І С Т

стор.

В С Т У П

Будь-який технологічний процес може бути охарактеризований певним числом фактором або вхідних параметрів, які в різноманітній мірі впливають на вихідні параметри, тобто на кількість продукту, його якісні або кількісні характеристики, що одержуються за ходом процесу.

Метою дослідження часто є встановлення кількісної залежності вихідного параметру будь-якого процесу від одного або групи вхідних параметрів в умовах коливання значень вхідних і вихідних параметрів, яка зумовлена впливом випадкових і в більшості своїй факторів, що не піддаються обліку.

Якщо взаємозв’язок між двома змінними величинами описується функцією $y=f(x)$, тоді в математичному аналізі ця залежність зветься функціональною. Це значить, що в відповідності з видом функції кожному значенню незалежної змінної x відповідає одно або декілька цілком певних значень залежної змінної y.

При вивченні взаємного впливу або зв’язку випадкових величин, котрими є практично всі параметри, які оцінюють в дослідницькій практиці, спостерігається інший вид зв’язку. Особливість його полягає в тому, що одному значенню змінної x може відповідати деяка сукупність значень залежної змінної y. Поява такої сукупності значень залежної змінної y викликане впливом множини побічних факторов, які діють одночасно або послідовно за різними напрямками. В цьому випадку зв’язок між змінними x та y на відміну від функціональної набуває статистичний характер і зветься кореляційним. Кореляційний зв’язок займає проміжне положення між суворо функціональною залежністю і повною відсутністю її між змінними.

Зміщення кореляційної залежності в ту або іншу сторону зумовлено “конкурентним” впливом двох складових. Одна з них (стохастична) визначається об’єктивно діючими фізичними і технологічними зв’язками між змінними. Друга складова (випадкова) є результатом впливу багатьох неврахованих факторів. Домінування першої складової зсуває кореляційну залежність в сторону функціонального зв’язку, а другої - в сторону повної незалежності випадкових величин.

1 ЛІНІЙНА РЕГРЕСІЯ. МЕТОД НАЙМЕНШИХ КВАДРАТІВ

Якщо між незалежною (вхідною) величиною x і залежною (вихідною) величиною y є або припускається кореляціний зв’язок, тоді її можна оцінити і дослідити за допомогою методу регресійного аналізу. Найпростіша і надто розповсюджена залежність між величинами x і y є лінійна регресія, на основі якої можна оцінювати лінійний або парний кореляційний зв’язок між цими величинами. Задача знаходження вибіркового рівняння регресії і наступної перевірки значущості його коефіцієнтів вирішується методами регресійного аналізу. Оцінка тісноти або сили зв’язку між величинами x і y здійснюється методами кореляційного аналізу. Математичний апарат регрегресійного і кореляційного аналізу в значній мірі містить спільні елементи.

Розглянемо лінійну регресію від одного параметру. Нехай для довільного фіксованого значення x отримано декілька значень. Припускається, що величина y розподілена нормально з математичним очікуванням:

$$m_y = b_0^* + b_1^* \cdot x \tag{1}$$

і дисперсією σ_y^2, яка не залежить від x. З виразу (1) слідує, що випадкова величина y в середньому лінійно залежить від фіксованого значення x, а параметри b_0^*, b_1^* та σ_y^2 є невідомими параметрами генеральної сукупності.

Для оцінки цих невідомих величин за вибіркою об’ємом n спряжених пар значень $x_1, y_1; x_2, y_2; \ldots; x_n, y_n$ у декартовій системі координат можна побудувати кореляційне поле, яке містить n точок. Розташування точок на кореляційному полі в загалі виявляється невипадковим і підкоряється певній залежності. Якщо нанести на поле середні значення $\overline{y_i}$, які відповідають всім значенням змінної x_i в інтервалах, що обмежені вертикальними лініями координатної сітки, тоді залежність y від x може стати більш очевидною. Ламана лінія, що з’єднує точки $\overline{y_i}$, які віднесені до середин інтервалів $x_{сер\,i}$, називається емпіричною лінією регресії. Із збільшенням кількості дослідів ламана лінія буде зглажуватися і, звільнюючись від випадкових зигзагів, наближатися до деякої граничної лінії - теоретичної лінії регресії. В загальному випадку форма лінії регресії визначається характером зв'язку між x і y.

Для лінійної залежності лінія регресії задається рівнянням прямої лінії:

$$y = \beta_0 + \beta_1 \cdot x, \tag{2}$$

яка повинна проходити максимально близько до точок кореляційного поля. Ця вимога звичайно реалізується застосуванням методу найменших квадратів і зводиться до того, щоб відстань по вертикалі між дослідними точками з координатами x_i, y_i і відповідними точками, які лежать на шуканій лінії регресії, було мінімальним. Цю умову можна записати у вигляді:

$$\sum_{i=1}^{n} \left[y_i - (\beta_0 + \beta_1 \cdot x_i) \right]^2 = min \tag{3}$$

Взявши частинні похідні рівняння (3) за коефіцієнтуми β_0 і β_1 та дорівнюючи їх нулю, знаходимо рівняння для оцінок b_0 і b_1 невідомих параметрів β_0 і β_1:

$$\sum_{i=1}^{n} (y_i - b_0 - b_1 \cdot x_i) = 0, \tag{4}$$

звідки

$$n \cdot b_0 + b_1 \cdot \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \tag{5}$$

та

$$\sum_{i=1}^{n} x_i \cdot (y_i - b_0 - b_1 \cdot x_i) = 0, \tag{6}$$

звідки

$$b_o \cdot \sum_{i=1}^{n} x_i + b_1 \cdot \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} (y_i \cdot x_i) \tag{7}$$

Оскільки

$$\frac{1}{n} \cdot \sum_{i=1}^{n} x_i = \bar{x} \quad \text{та} \quad \frac{1}{n} \cdot \sum_{i=1}^{n} y_i = \bar{y}, \tag{8}$$

тоді з рівнянь (5) і (7) слідує

$$b_o = \bar{y} - b_1 \cdot \bar{x}, \tag{9}$$

$$b_1 = \frac{\sum_{i=1}^{n}(x_i \cdot y_i) - \left(\sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} x_i \;/\; n\right)}{\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2 \;/\; n} = \frac{\sum_{i=1}^{n}\left[\left(x_i - \bar{x}\right)\cdot\left(y_i - \bar{y}\right)\right]}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \tag{10}$$

Враховуючи співвідношення (9), вібіркове рівняння лінійної регресії y відносно x можна записати у вигляді:

$$y = b_0 + b_1 \cdot x = \bar{y} + b_1 \cdot \left(x_i - \bar{x}\right) \tag{11}$$

2 ВИБІРКОВИЙ КОЕФІЦІЄНТ КОРЕЛЯЦІЇ

Кількісною мірою, яка враховує закономірну (стохастичну) частку колебань y_i відносно середньої $\bar{y}$ під впливом x_i, є коефіцієнт кореляції. Вибірковій коефіцієнт кореляції обчислюють за формулою:

$$r = \frac{1}{(n-1)\cdot S_x \cdot S_y} \cdot \sum_{i=1}^{n}\left[\left(x_i - \bar{x}\right)\cdot\left(y_i - \bar{y}\right)\right] \tag{12}$$

де S_x та S_y - вибіркові середні квадратичні відхилення:

$$S_x = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}\;, \tag{13}$$

$$S_y = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \tag{14}$$

Коефіцієнт кореляції не може бути використаним для оцінки технологічної важливості процесу. Його величина вказує тільки на тісноту зв'язку між змінними, а знак "–" на характер впливу. Значення коєфіцієнту кореляції може знаходитися у межах $-1 \le r \le 1$. Якщо $r < 0$, тоді збільшення x викликає зменьшення y; при $r > 0$ спостерігається зворотня закономірність. Якщо $|r| = 1$, тоді зв'язок є лінійним функціональним, якщо ж $|r| = 0$, тоді кореляційного зв'язку між x і y нема або він нелінійний. Коефіцієнт кореляції однаково "реагує" на розкид експериментальних точок відносно прямої регресії і на криволінійність залежності при малому розкиду точок на кореляційному полі. Тому візуальний аналіз кореляційного поля може дати корисну інформацію щодо пояснення причини одержання малого значення коефіцієнту кореляції.

Якщо вираз (12) перетворити до вигляду

$$\sum_{i=1}^{n}\left[\left(x_i-\bar{x}\right)\cdot\left(y_i-\bar{y}\right)\right]=r\cdot S_x\cdot S_y\cdot\left(n-1\right) \tag{15}$$

і підставити його до формули (10), тоді отримаємо

$$b_1=r\cdot S_x\cdot S_y\cdot\left(n-1\right)/\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2=r\cdot S_x\cdot S_y\;/\;S_x^2=r\cdot S_y\;/\;S_x \tag{16}$$

З виразу (16) видний безпосередній зв'язок величин r і b_1, знаки котрих завжди співпадають. Вирази (12), (16) і (9) складають "суміщений" розрахунковий апарат для рішення переважної більшості практичних задач, в яких важливо знаходження тісноти і виду зв'язку між змінними x і y.

Коефіцієнт кореляції звичайно розраховують за обмеженою кількістю даних - вибірки з генеральної сукупності, внаслідок чого він завжди містить помилку. Тому необхідна перевірка гіпотези о його статистичній значущості, тобто відзнаки від нуля генерального коефіцієнту r^*. Для перевірки нуль-гіпотези $H_0 : r^* = 0$ застосовують t-відношення:

$$t_r < t(\alpha, f), \tag{17}$$

де t_r - розрахункове значення коефіцієнту Стьюдента, знаходиться з наступного рівняння:

$$t_r=\left|r\right|\cdot\frac{\sqrt{n-2}}{\sqrt{1-r^2}}; \tag{18}$$

$t(\alpha, f)$ - табличне значення коефіцієнту Стьюдента при заданому рівні значущості α і ступеню свободи f, які знаходяться з наступних виразів:

$$f=n-2; \tag{19}$$

$$\alpha=1-\beta, \tag{20}$$

де β - довірча імовірність, n - кількість дослідів.

Якщо $t_r > t(\alpha, f)$ при заданому рівні значущості, тоді нульова гіпотеза відхиляється і приймається альтернативна $H_1 : r^* \neq 0$, тобто r значущо відрізняється від нуля (H_0 - основна нуль-гіпотеза, H_1 - альтернативна або конкурентна гіпотеза, яка наводиться для зіставлення з основною для випробування останньої). Так, для α=0,01 (тобто β=0,99) і $f=\infty$ (див. додаток 1) знаходимо $t(\alpha, f) = 2{,}58$. Таким чином, при $t_r > 2{,}58$ зв'язок між факторами вважається не випадковим.

3 ПЕРЕВІРКА ЗНАЧУЩОСТІ КОЕФІЦІЄНТІВ РІВНЯННЯ РЕГРЕСІЇ ТА ЙОГО АДЕКВАТНОСТІ

Невід'ємним елементом регресійного аналізу є статистична перевірка значущості знайдених коефіцієнтів регресії. Оцінку значущості коефіцієнтів рівняння виконують за крітерієм Стьюдента (**t**-крітерієм). При цьому перевіряється нуль-гіпотеза $H_0:\ b_j^* = 0$, тобто значення **j**-го коефіцієнту регресії генеральної сукупності при заданому рівні значущості α невідрізнимо від нуля. Якщо умова

$$t_{b_j} \le t(\alpha, f) \tag{21}$$

виконується, тоді нульова гіпотеза приймається. При недотриманні умови (21) приймається альтернативна гіпотеза $H_1:\ b_j^* \ne 0$. У випадку прийняття нуль-гіпотези незначущий коефіцієнт виключається з рівняння регресії, а величини коефіцієнтів, що залишилися, знаходять ще раз, так як між ними існує кореляційний зв'язок (9).
В умові (21) $f = n - k$ - число ступенів свободи; k - число ознак, що ураховуються в рівнянні регресії. t_{b_j} - розрахункове значення коефіцієнту Стьюдента для **j**-го коефіцієнту регресії b_j, знаходиться з наступного виразу:

$$t_{b_j} = \frac{|b_j|}{S_{b_j}}, \tag{22}$$

де S_{b_j} - середнє квадратичне відхилення **j**-го коефіцієнту.

Середні квадратичні відхилення S_{b_j} коефіцієнтів лінійної регресії для перевірки умови (21) знаходять за формулами:

$$S_{b_1} = S_{ост} \,/\, \sqrt{(n-1) \cdot S_x^2}\ ; \tag{23}$$

$$S_{b_0} = S_{ocm} \,/\, \sqrt{\frac{1}{n} + \frac{x^2}{(n-1) \cdot S_x^2}}\,, \tag{24}$$

де S_{ocm} - корінь квадратний із остаточної дисперсії, яку обчислюють за формулою:

$$S_{ocm}^2 = \frac{1}{n-k} \cdot \sum_{i=1}^{n} (y_i - \hat{y}_i)^2\,, \tag{25}$$

де $\hat{y}_i$ - величини, які обчислені за рівнянням регресії; k - число ознак, що урахову-ються, у рівнянні регресії (для лінійної регресії $k=2$); $f=n-k$ - число ступенів свободи. Якщо коефіцієнт кореляції r вже обчислений, тоді при виконанні практичних розрахунків зручно використовувати зв'язок між лінійною кореляцією і лінійною регресією. У цьому випадку для знаходження остаточної дисперсії можна використовувати формулу:

$$S_{ocm}^2 = \frac{n-1}{n-2} \cdot S_y^2 \cdot \left(1 - r^2\right) \tag{26}$$

Іншим важливим елементом регресійного аналізу є перевірка адекватності рівняння регресії за крітерієм Фішера (F-крітерієм). У цьому випадку перевіряється нуль-гіпотеза $H_0:\ \sigma_{a\partial}^2 = \sigma_{\text{від}}^2$, тобто припускається, що генеральні дисперсії адекватності $\sigma_{a\partial}^2$ та відтворюємості $\sigma_{\text{від}}^2$ рівні. Оскільки перевірка здійснюється шляхом порівняння вибіркових дисперсій, то нуль-гіпотеза приймається при виконанні умови:

$$F_p < F_{1-\alpha}(f_1, f_2)\,, \tag{27}$$

де F_p - розрахункове значення крітерия Фішера, знаходиться з виразу:

$$F_p = S_{a\partial}^2 \,/\, S_{\text{від}}^2\,, \tag{28}$$

де $S_{a\partial}^2$ - вибіркова дисперсія адекватності; $S_{\text{від}}^2$ - вибіркова дисперсія відтворюємості; $f_1 = f_{a\partial}$ - число ступенів свободи для $S_{a\partial}^2$; $f_2 = f_{\text{від}}$ - число ступенів свободи для $S_{\text{від}}^2$.

При повторенні (дублюванні) кожного з n дослідів m разів дисперсії адекватності і відтворюємості обчислюють за формулами:

$$S_{ад}^2 = m \cdot \sum_{i=1}^{n} (\bar{y}_i - \hat{y}_i)^2 \;/\; (n-k), \tag{29}$$

$$S_{від}^2 = \sum_{i=1}^{n} \sum_{u=1}^{m} (y_{iu} - \bar{y}_i)^2 \;/\; \left[n \cdot (m-1) \right], \tag{30}$$

де n - об'єм вибірки; m - число дублюючих дослідів; k - число коефіцієнтів в рівнянні регресії; $\hat{y}_i$ - значення, яке обчислено за рівнянням регресії для x_i; $f_1 = n - k$; $f_2 = n \cdot (m - 1)$.

У випадку неможливості проведення дублюючих дослідів та визначення дисперсії відтворюємості замість співвідношення (27) для оцінки адекватності рівняння регресії використовують "зворотне" співвідношення дисперсій:

$$F_p > F_{1-\alpha}(f_1, f_2), \tag{31}$$

де $f_1 = n - 1$; $f_2 = n - k$.

У цьому випадку розрахункове значення крітерія Фішера F_p знаходять з виразу:

$$F_p = S_y^2 \;/\; S_{ост}^2, \tag{32}$$

У виразі (32) остаточну дисперсію $S_{ост}^2$ знаходять за формулами (25) і (26), а дисперсію відносно середнього S_y^2 - за формулою (14). Ефективність рівняння регресії тим вище, чим більше F_p перевищує $F_{1-\alpha}(f_1, f_2)$.

Приклад 1. З легованого високопрочного чавуну відлита партія валів, кількість котрих склала n=42, з диаметром бочки 400...600 мм. В якості основного легуючого елемента для регулювання твердості робочого шару валів застосовували нікель. Результати вимірювань твердості робочого шару (у одиницях Шора) на глибині 5 мм від литої поверхні і вміст нікелю у чавуні цих валів наведені у таблиці 1 і на рис.1.

Необхідно при рівні значущості $\alpha = 0,01$:

- оцінити тісноту лінійного кореляційного зв'язку між вмістом нікеля (x) і твердістю чавуну (y) та перевірити значущість коефіцієнту кореляції;

- визначити коефіцієнти рівняння лінійної регресії і перевірити їхню статистичну значущість;
- оцінити якість апроксимації експериментальних даних отриманим лінійним рівнянням регресії.

Рішення:

Будуємо кореляційне поле за експеріментальними точками і проводимо емпіричну лінію регресії (рис.1).

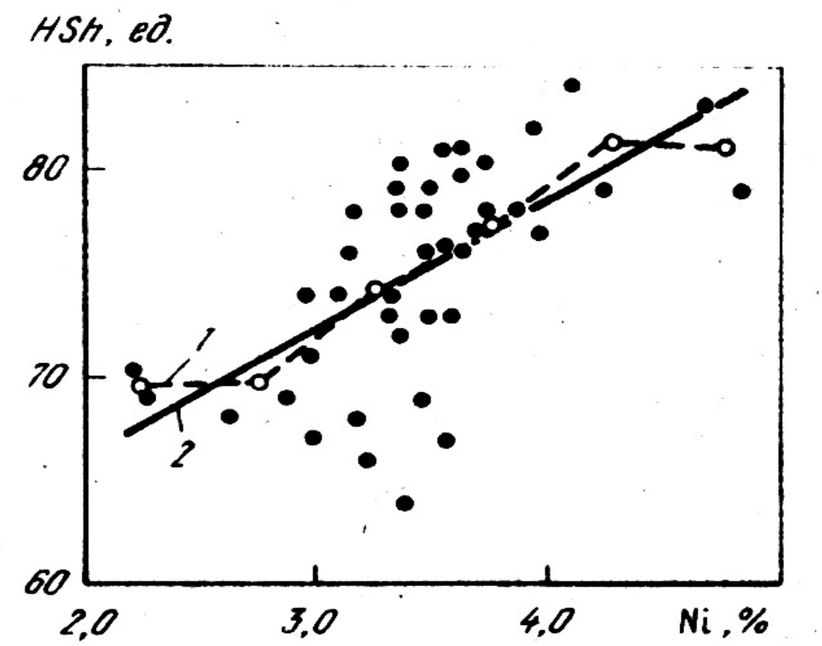

Рисунок 1 ***–*** Кореляційне поле залежності твердість чавуну-вміст нікелю (***HSh-Ni***) з емпіричною (1) і теоретичною (2) лініями регресії

Виконавши попередні розрахунки в відповідності за формою табл.1, за формулами (8), (13), (14), (12), знаходимо:

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = \frac{1}{42} \cdot 144{,}06 = 3{,}43; \quad \bar{y} = \frac{1}{n} \cdot \sum_{i=1}^{n} y_i = \frac{1}{42} \cdot 3150{,}84 = 75{,}02;$$

$$S_x = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2} = \sqrt{\frac{1}{42-1} \cdot 10{,}4931} = 0{,}506; \quad S_x^2 = 0{,}256;$$

$$S_y = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2} = \sqrt{\frac{1}{42-1} \cdot 1063} = 5{,}092; \quad S_y^2 = 25{,}93;$$

$$r = \frac{1}{(n-1)\cdot S_x \cdot S_y} \cdot \sum_{i=1}^{n}\left[\left(x_i - \overline{x}\right)\cdot\left(y_i - \overline{y}\right)\right] = \frac{1}{(42-1)\cdot 0{,}506 \cdot 5{,}092} \cdot 65{,}57 = 0{,}62$$

Перевіряємо статистичну значущість коефіцієнту кореляції, формулюючи нуль-гіпотезу $H_0 : r^* = 0$, де r^* - генеральний коефіцієнт кореляції. За формулою (18) знаходимо:

$$t_r = |r| \cdot \frac{\sqrt{n-1}}{\sqrt{1-r^2}} = |0{,}62| \cdot \frac{\sqrt{42-2}}{\sqrt{1-0{,}62^2}} = 5$$

Для $\alpha = 0{,}01$ і $f = n - 2 = 42 - 2 = 40$ (Додаток 1) знаходимо $t(\alpha, f) = 2{,}71$. Оскільки $t_r > t(\alpha, f)$, гіпотеза H_0 повинна бути отвергнута, так як умова (17) не виконується. Вибірка з 42-х пар спостережень взята з генеральної сукупності з коефіцієнтом кореляції $r^* \neq 0$. Отже, випадкові величини x (концентрація нікеля) і y (твердість) зв’язані лінійною залежністю.

Розраховуємо за формулою (16) коефіцієнт регресії b_1:

$$b_1 = r \cdot S_y \; / \; S_x = 0{,}62 \cdot 5{,}092 \; / \; 0{,}506 = 6{,}24$$

За формулою (11) записуємо рівняння лінійної регресії:

$$y = b_0 + b_1 \cdot x = \overline{y} + b_1 \cdot \left(x - \overline{x}\right) = 75{,}02 + 6{,}24 \cdot (x - 3{,}43) = 53{,}6 + 6{,}24 \cdot x,$$

де $53{,}6 = b_0$.

Цьому рівнянню відповідає пряма лінія, яка нанесена на поле кореляції за двома довільними точками (рис.1). Наприклад, задавши у рівнянні регресії $x_1=2$ і $x_2=4$, знаходимо $y_1=66{,}1$ і $y_2=78{,}6$. Слід мати на увазі, що користуватися отриманим рівнянням регресії можна тільки у досліджених межах зміни концентрацій нікеля, тобто 2,2 % ≤ x ≤ 4,8 %.

Таблиця 1 – Вхідні дані і результати розрахунків до прикладу 1

Дос-лід, i	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	y_i	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})^2 \times$ $\times (y_i - \bar{y})^2$
1	3,68	0,25	0,0625	78	3	9	0,75
2	3,70	0,27	0,0729	77	2	4	0,54
3	3,44	0,01	0,0001	75	0	0	0,00
4	3,48	0,05	0,0025	79	4	16	0,20
5	2,16	-1,27	1,6129	70	-5	25	6,35
6	2,25	-1,18	1,3924	69	-6	36	7,08
7	3,34	-0,09	0,0081	78	3	9	-0,27
8	3,36	-0,07	0,0049	79	4	16	-0,028
9	2,95	-0,48	0,2304	74	-1	1	0,48
10	2,61	-0,82	0,6724	68	-7	49	5,74
11	3,60	0,17	0,0289	80	5	25	0,85
12	3,52	0,09	0,0081	73	-2	4	-0,18
13	3,94	0,51	0,2601	77	2	4	1,02
14	3,00	-0,43	0,1849	67	-8	64	3,44
15	3,90	0,47	0,2209	82	7	49	3,29
16	3,60	0,17	0,0289	81	6	36	1,02
17	3,36	-0,07	0,0049	72	-3	9	0,21
18	3,32	-0,11	0,0121	74	-1	1	0,11
19	3,53	0,10	0,0100	81	6	36	0,60
20	3,44	0,01	0,0001	69	-6	36	-0,06
21	3,56	0,13	0,0169	73	-2	4	-0,26
22	4,64	1,21	1,4641	83	8	64	0,68
23	3,50	0,07	0,0049	76	1	1	0,07
24	3,47	0,04	0,0016	76	1	1	0,04

Закінчення таблиці 1

Дос-лід, i	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	y_i	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})^2 \times$ $\times (y_i - \bar{y})^2$
25	3,60	0,17	0,0289	76	1	1	0,17
26	3,18	-0,25	0,0625	68	-7	49	1,75
27	3,70	0,27	0,0729	80	5	25	1,35
28	4,80	1,37	1,8769	79	4	16	5,48
29	3,07	-0,36	0,1296	74	-1	1	0,36
30	3,84	0,41	0,1681	78	3	9	1,23
31	2,84	-0,59	0,3481	69	-6	36	3,54
32	3,30	-0,13	0,0169	73	-2	4	0,26
33	3,35	-0,08	0,0064	80	5	25	-0,40
34	3,12	-0,31	0,0961	76	1	1	-0,31
35	4,22	0,79	0,6241	79	4	16	3,16
36	3,00	-0,43	0,1849	71	-4	16	1,72
37	3,55	0,12	0,0144	67	-8	64	-0,96
38	3,37	-0,06	0,0036	64	-11	121	0,66
39	3,20	-0,23	0,0529	66	-9	81	2,07
40	3,45	0,02	0,0004	78	3	9	0,06
41	3,15	-0,28	0,0784	78	3	9	-0,84
42	4,08	0,65	0,4225	84	9	81	5,85
Σ	144,06	0,11	10,4931	3150,84	1	1063	65,57
$\bar{x}, \bar{y}$	3,43	-	-	75,02	-	-	-

Перевіряємо статистичну значущість коефіцієнту регресії b_1, тобто гіпотезу $H_0 :\ b_1^* = 0$. Для цього обчислюємо за формулою (26) остаточну дисперсію:

$$S_{ocm}^2 = \frac{n-1}{n-2} \cdot S_y^2 \cdot \left(1 - r^2\right) = \frac{42-1}{42-2} \cdot 25{,}93 \cdot (1 - 0{,}62^2) = 16{,}36 ,$$

звідки $S_{ocm} = 4{,}04$. При b_1 = 6,24; S_x = 0,506 і S_{ocm} = 4,04 за формулами (23) і (22) знаходимо

$$S_{b_1} = S_{ocm} \ / \ \sqrt{(n-1) \cdot S_x^2} = 4{,}04 \ / \ \sqrt{(42-1) \cdot 0{,}256} = 1{,}247 ;$$

$$t_{b_1} = \frac{|b_1|}{S_{b1}} = \frac{|6{,}24|}{1{,}247} = 5$$

Для $\alpha = 0{,}01$ і $f = n - 2 = 42 - 2 = 40$ (Додаток 1) знаходимо $t(\alpha, f) = 2{,}71$. Так як $t_{b_1} > t(\alpha, f)$, тоді нуль-гіпотезу H_0 отвергаємо. Величина y (твердість чавуну) значущо залежить від величини x (вмісту нікеля).

Перевіряємо нуль-гіпотезу $H_0 :\ b_0^* = 0$ (про рівність нулю константи b_0 в отриманому рівнянні регресії). При $n = 42$; $\bar{x} = 3{,}43$; $S_x = 0{,}506$; $S_{ocm} = 4{,}04$; $b_0 = 53{,}6$ за формулами (24) і (22) знаходимо:

$$S_{b_0} = S_{ocm} \ / \ \sqrt{\frac{1}{n} + \frac{x^2}{(n-1) \cdot S_x^2}} = 4{,}04 \ / \ \sqrt{\frac{1}{42} + \frac{3{,}43^2}{(42-1) \cdot 0{,}506^2}} = 4{,}32$$

$$t_{b_0} = \frac{|b_0|}{S_{b0}} = \frac{|53{,}6|}{4{,}32} = 12{,}41$$

Для $\alpha = 0{,}01$ і $f = n - 2 = 42 - 2 = 40$ (Додаток 1) табличне значення коефіцієнту Стьюдента рівно $t(\alpha, f) = 2{,}71$. Оскільки $t_{b_0} > t(\alpha, f)$, нульова гіпотеза H_0 отвергається і приймаємо, що b_0 значущо відрізняется від нуля.

Перевіряємо адекватность отриманого рівняння лінійної регресії за умовою (31). Використовуючи обчислені раніш за формулами (26) і (14) значення $S_{ocm}^2 = 16{,}36$ і $S_y^2 = 25{,}93$. За формулою (27) знаходимо розрахункове значення F-критерія:

$$F_p = S_y^2 \ / \ S_{ocm}^2 = 25{,}93 \ / \ 16{,}36 = 1{,}58$$

З Додатку 3 для $\alpha = 0{,}01$; $f_1 = n - 1 = 42 - 1 = 41$ і $f_2 = n - k = 42 - 2 = 40$ знаходимо квантіль розподілу Фішера $F_{1-\alpha}(f_1, f_2)$:

$$F_{1-0{,}01}(41, 40) \approx F_{1-0{,}01}(\infty, 40) = 1{,}8$$

Так як $F_p < F_{1-\alpha}(f_1, f_2)$, тоді отримане рівняння регресії не дає необхідної якості апроксимації при рівні значущості $\alpha = 0{,}01$. Виконаємо перевірку для рівня значущості $\alpha = 0{,}05$. З Додатку 2 знаходимо:

$$F_{1-0{,}05}(41, 40) \approx F_{1-0{,}05}(\infty, 40) = 1{,}5$$

Так як $F_p > F_{1-\alpha}(f_1, f_2)$, тоді при рівні значущості $\alpha = 0{,}05$ можна рахувати, що отримане рівняння регресії задовільно апроксимірує характер взаємодії факторів x і y.

4 ВХІДНІ ДАНІ ДЛЯ СТАТИСТИЧНОЇ ОБРОБКИ

Використовуючи результати розрахунків охолодження зливка, побудувати емпіричну лінію регресії залежності температури в заданій точці з координатами i та j на поверхні зливка від часу процесу. Використовуючи метод найменших квадратів, побудувати теоретичну лінію регресії, яка носить лінійний характер. Виконати статистичну перевірку значущості коефіцієнту кореляції і коефіцієнтів рівняння регресії, а також його адекватності.

Варіант 1: $i=m$; $j=n$.

Варіант 2: $i=1$; $j=n$.

Варіант 3: $i=m$; $j=n$.

Варіант 4: $i=m$; $j=1$.

Варіант 5: $i=m$; $j=n$.

Варіант 6: $i=m$; $j=1$.

Варіант 7: $i=m$; $j=1$.

Варіант 8: $i=1$; $j=3$.

Варіант 9: $i=m$; $j=1$.

Варіант 10: ***i=1***; ***j=5***.

Варіант 11: ***i=m***; ***j=1***.

Варіант 12: ***i=1***; ***j=3***.

Варіант 13: ***i=m***; ***j=n***.

Варіант 14: ***i=3***; ***j=n***.

Варіант 15: ***i=5***; ***j=1***.

Варіант 16: ***i=6***; ***j=n***.

Варіант 17: ***i=1***; ***j=3***.

Варіант 18: ***i=m***; ***j=n***.

Варіант 19: ***i=m***; ***j=1***.

Варіант 20: ***i=1***; ***j=n***.

Варіант 21: ***i=1***; ***j=3***.

Варіант 22: ***i=m***; ***j=1***.

Варіант 23: ***i=1***; ***j=4***.

Варіант 24: ***i=1***; ***j=n***.

Варіант 25: ***i=4***; ***j=1***.

Варіант 26: ***i=1***; ***j=2***.

Варіант 27: ***i=m***; ***j=n***.

Варіант 28: ***i=1***; ***j=2***.

Варіант 29: ***i=m***; ***j=2***.

Варіант 30: ***i=1***; ***j=2***.

Звіт по статистичній обробці експериментальних даних повинен бути набраний і роздрукований у середовищі текстового редактору Microsoft Word 8.0 (або 7.0) та містити у собі:

1. Умову завдання.
2. Сформульовану постановку задачі для свого варіанту.
3. Опис алгоритму розрахунків.
4. Листінг програми на алгоритмічній мові Turbo Pascal.
5. Результати розрахунків.

6. За експериментальними даними результатами розрахунку у середовищі табличного процесору Microsoft Excel 8.0 (або 7.0) побудувати кореляційне поле, емпіричну і теоретичну лінії регресії.

7. Зробити висновки про значущість коефіцієнтів кореляції і рівняння регресії, а також про характер взаємодії факторів ***x*** та ***y***.

ЛІТЕРАТУРА, ЩО РЕКОМЕНДУЄТЬСЯ

1. Белай Г.Е., Дембовский.В.В., Соценко О.В. Организация металлургического эксперимента.- М: Металлургия, 1993.- 256 с.

Додаток 1

Табличне значення ***t***-критерія (критерія Стьюдента)

для довірчої імовірності β і числа ступенів свободи f

f \ β	**0,80**	**0,90**	**0,95**	**0,99**	**0,999**
1	3,08	6,31	12,71	63,66	636,62
2	1,89	2,92	4,30	9,93	31,60
3	1,64	2,35	3,18	5,84	12,92
4	1,53	2,13	2,78	4,60	8,61
5	1,47	2,02	2,57	4,03	6,87
6	1,44	1,94	2,45	3,71	5,96
7	1,42	1,90	2,37	3,50	5,41
8	1,40	1,86	2,31	3,36	5,04
9	1,38	1,83	2,26	3,25	4,78
10	1,37	1,81	2,23	3,17	4,59
11	1,36	1,80	2,20	3,11	4,44
12	1,36	1,78	2,18	3,05	4,32
13	1,35	1,77	2,16	3,01	4,22
14	1,34	1,76	2,15	2,98	4,14
15	1,34	1,75	2,13	2,95	4,07
16	1,34	1,75	2,12	2,92	4,02
17	1,33	1,74	2,11	2,90	3,97
18	1,33	1,73	2,10	2,88	3,92
19	1,33	1,73	2,09	2,86	3,88
20	1,33	1,73	2,09	2,85	3,85
25	1,32	1,71	2,06	2,79	3,73
30	1,31	1,70	2,04	2,75	3,65
40	1,30	1,68	2,02	2,71	3,55
60	1,30	1,67	2,00	2,66	3,46
120	1,29	1,66	1,98	2,62	3,37
∞	1,28	1,64	1,96	2,58	3,30

Додаток 2

Значення $\boldsymbol{F}$-крітерія (крітерія Фішера) для довірчої імовірності β=0,95 при відповідних числа ступенів свободи числівника f_1 і знаменника f_2

f_2	f_1								
	1	**2**	**3**	**4**	**5**	**6**	**8**	**12**	**∞**
1	161	200	216	225	230	234	239	244	254
2	18,5	19,2	19,3	19,3	19,3	19,4	19,4	19,4	19,5
3	10,1	9,6	9,1	9,0	8,9	8,7	8,8	8,74	8,53
4	7,7	6,6	6,6	6,4	6,3	6,2	6,0	5,91	5,63
6	6,0	5,1	4,8	4,5	4,4	4,3	4,2	4,00	3,67
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	2,93
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,54
12	4,75	3,88	3,49	3,26	3,14	3,00	2,85	2,69	2,30
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,13
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,01
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	1,92
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	1,84
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	1,78
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,73
26	4,23	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,69
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,65
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,51
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,00

Додаток 3

Значення ***F***-критерія (крітерія Фішера) для довірчої імовірності β=0,99 при відповідних числа ступенів свободи числівника f_1 і знаменника f_2

f_2	f_1								
	1	**2**	**3**	**4**	**5**	**6**	**8**	**12**	**∞**
1	4052	4999	5403	5625	5764	5859	5981	6106	6366
2	98,5	99,1	99,2	99,3	99,3	99,3	99,4	99,4	99,5
3	23,1	30,8	29,5	28,7	28,2	27,9	27,5	27,1	26,1
4	21,2	18,0	16,7	15,3	15,5	15,2	14,8	14,4	13,5
6	13,7	10,9	9,78	9,15	8,75	8,47	8,10	7,52	9,02
8	11,3	8,65	7,59	7,01	6,63	6,37	6,03	5,67	4,86
10	10,0	7,56	6,55	5,99	5,64	5,39	5,06	4,85	3,91
12	9,33	6,93	5,95	5,41	5,06	4,82	4,50	4,16	3,36
14	8,86	6,51	5,56	5,03	4,69	4,46	4,14	3,80	3,00
16	8,53	6,23	5,29	4,77	4,44	4,29	3,89	3,55	2,75
18	8,28	6,01	5,09	4,58	4,25	4,01	3,71	3,37	2,57
20	8,10	5,85	4,94	4,43	4,10	3,87	3,56	3,23	2,42
22	7,94	5,72	4,82	4,31	3,99	3,76	3,45	3,12	2,31
24	7,82	5,61	4,72	4,22	3,90	3,67	3,36	3,03	2,21
26	7,72	5,53	4,64	4,14	3,82	3,59	3,29	2,96	2,13
28	7,64	5,45	4,57	4,07	3,76	3,53	3,23	2,90	2,06
30	7,56	5,39	4,51	4,02	3,70	3,47	3,17	2,84	2,01
40	7,31	5,18	4,31	3,83	3,51	3,29	2,99	2,66	1,81
∞	6,64	4,60	3,78	3,32	3,02	2,80	2,41	2,18	1,09

НАВЧАЛЬНЕ ВИДАННЯ

Регресійна модель з однією вхідною змінною для статистичної обробки даних на ПЕОМ. Методичні вказівки до практичної роботи з дисципліни «Моделювання технологічних та фізичних процесів» освітньо-наукової програми вищої освіти (підготовка докторів філософії) зі спеціальності 136 "Металургія".

Укладач: Пантейков Сергій Петрович

Підписано до друку __________ 2017 р.
Формат 60х84 $^{1}/_{16}$. Обсяг ______ др. арк.
Тираж _____ прим. Замовлення № ___
51918, м. Кам'янське,
вул. Дніпробудівська, 2