СТАТИСТИЧНІ МЕТОДИ В ХЕМІЇ, ХЕМІЧНІЙ ТЕХНОЛОҐІЇ ТА МАТЕРІАЛОЗНАВСТВІ

УДК 620.24/27:543/545

Г.О. Сіренко, М.Б Складанюк, М.І. Мартинюк, І.В. Говдяк

Статистичні методи в хемії та хемічній технолоґії: 2. Кореляційна аналіза (теорія)

[1]*Прикарпатський національний університет імені Василя Стефаника, вул. Шевченка, 57, м. Івано-Франківськ, 76018, Україна*

У статті розглядаються теоретичні основи кореляційної аналізи. Дано поняття про детерміновану (функціональну) та стохастичну (кореляційну) залежности між двома неперервними випадковими величинами. Формулюється завдання кореляційної аналізи. Приведені формули для: розрахунку генерального та вибіркового коефіцієнтів кореляції, перевірки значущости коефіцієнта кореляції за критеріями Стюдента і Фішера та створення довірчих інтервалів для коефіцієнта кореляції. Дана методика порівняння двох генеральних коефіцієнтів кореляції.

Ключові слова: функціональна залежність, кореляційна залежність. Кореляція, лінійний та нелінійний зв'язок, довірчі інтервали, значущість коефіцієта кореляції.

H.O.Sirenko, M.B. Skladanyuk, M.I. Martynyuk

Mathematical methods in chemistry and chemical technology: dispersion analysis

[1]*Vasyl Stefanyk Precarpathian National University,57, Shevchenko Str., Ivano-Frankivsk, 76018, Ukraine*

A manuscript that describes mathematical methods in chemistry and chemical technology, namery the teory of dispersion analysis. The material on the application of dispersion analysis in chemical technology has been worked out.

Keywords: variance, statistical equality, dispersion analysis

Стаття поступила до редакції 15.09.2018; прийнята до друку 05.12.2018.

Вступ

Актуальність теми. Метою більшости досліджень у хемії та хемічній технології є вирішення складних багаточинникових експериментальних завдань, які пов`язані зі встановленням надійних зв`язків між неперервними випадковими величинами, пошуком оптимальних рішень якости матеріалів та оптимальних умов проведення хеміко – технологічних процесів, розробкою раціональних конструкцій хемічного обладнання тощо.

1. Кореляційна аналіза (теорія). Означення. Розрахункові формули

1.1. Загальні означення. Постановка завдань дослідження

1. Кореляція [(ko + лат. relation) – сумісне + відношення; пізньолат. correlatio – співвідношення] - взаємозв'язок, взаємозалежність, взаємовідповідність; співвідношення понять, функцій, предметів, підприємств; почуттів, поведінки людей тощо [мат. Кореляція - ймовірнісна або статистична залежність між випадковими величинами]. Вперше (у 1800 – 1805 р.р.) у науковий обіг було введено означення «кореляція» французьким зоологом, анатомом, палеонтологом, історикм природничих наук Жоржем Кув'є [Cuvier (1769 –1832)]. Кореляційна аналіза тісно пов'язана з регресійною аналізою (регре́сія від лат. regressio – зведення, звертання, зворочення, що від лат. regredior – повертаюся; regressus – зворотній рух, відход).

2. Існує цілий клас завдань, пов'язаних із аналізою зв'язку між двома *(x~y)* чиількома змінними ($x_1, x_2, \ldots \sim y$), або ($x_1, x_2, \ldots \sim y_1, y_2, \ldots$).

3. Розрізняють два види математичних залежностей (зв'язків) однієї величини від іншої або кількох інших величин:

а) **детермінована** (від лат. determinatio – обмеження, визначення), функціональна; невипадкова; жорстка; чиста; строга; математична; регулярна залежність тощо, коли кожному значенню змінної (аргументу (***x***)) або кільком залежним величинам (аргументам $x_1, x_2, \ldots, x_j, \ldots x_n$) відповідає певне значення іншої величини (функції (***y***)) або певним значенням інших величин. **Такий зв'язок:**

$y = f(x)$ або $y = f(x_1, x_2, \ldots, x_n)$ (1)

носить назву **функціонального**, ...

Так, величина $S = \frac{gt^2}{2}$, де g = const, кожному значенню аргумента (***t***) відповідає певне детерміноване значення функції (***S***);

б) **випадкова**, коли між випадковими величинами чистого детермінованого зв'язку немає. Залежність між випадковими величинами зветься стохастичною (від дгр. stochasis – здогадка), випадковою, ймовірністною, недетермінованою, нежорсткою, нестрогою, нерегулярною, кореляційною тощо, коли кожному значенню аргумента з параметрами розподілу a_x, σ_x^2, σ_x. . . відповідає певний закон розподілу функції, тобто, коли на зміну однієї (або кількох) випадкових величин інша випадкова величина реагує зміною параметрів свого розподілу (a_y, σ^2_y, σ_y, . . .). Такий зв'язок носить назву стохастичного, ймовірного, кореляційного, . . . Стохастичний процес – це процес, характер зміни якого у часі не можна точно передбачити.

4. Стохастичні зв'язки існують, тому, що, окрім впливу основних чинників та їх рівнів, є:

• вплив випадкових неконтрольованих та некерованих чинників та їх рівнів на об'єкт дослідження або природний об'єкт спостереження;

• вплив випадкових контрольованих але некерованих чинників.

Окрім того, неможливо, практично не вдається зафіксувати рівні чинників, вони «плавають» біля певної середньої.

Все це впливає на функцію відгуку і вона теж має випадкові значення.

5. Стохастичний зв'язок має дві складові:

1) **власне стохастичну складову *St***, яка пов'язана із залежністю випадкових величин ***y*** і визначається дією залежних чинників;

2) **власне випадкову складову *Ra***, яка обумовлена дією індивідуальних випадкових чинників, які впливають на одну із випадкових величин.

При цьому:

а) якщо $St \neq 0$, $Ra = 0$ – функціональний зв'язок;

б) якщо $St = 0$, $Ra \neq 0$ (чи і $Ra = 0$) – незалежні величини;

в) якщо $St \neq 0$, $Ra \neq 0$ – стохастичний (кореляційний) зв'язок;

г) якщо одна величина ***y*** (змінна $\boldsymbol{a_y}$ *і* $\boldsymbol{\sigma^2_y}$) не реагує на зміну $\boldsymbol{a_x}$ другої величини ***x***, а реагує лише на зміну її дисперсії $\boldsymbol{\sigma^2_x}$, то такі величини є некорельовані.

6. Від співвідношення *St: Ra* залежить сила (щільність) зв'язку і це знаходження цього відношення величин є предметом кореляційної аналізи.

7. Тому, завдання кореляційної аналізи (разом із регре́сійною аналізою) полягає у тому, щоби встановити якісні показники математичної моделі та **дати відповідь на питання:**

• Наскільки лінійна чи нелінійна математична модель краще, добротніше, адекватно [адеква́тний (від лат. adaequātus – прирівняний, відповідний, цілком відповідний, тотожний)] описує залежність y~f(x), [чи $y = f(x_1, x_2, \ldots, x_i \ldots, x_n)$ тощо].

На рис. 1 показана апроксима́ція [(від лат. approximare – наближатися). мат. Заміна одних математичних об'єктів іншими, близькими до вихідних; приблизне вираження яких-небудь величин через інші, простіші величини] результатів експерименту чи спостережень лінійною та нелінійною функціями.

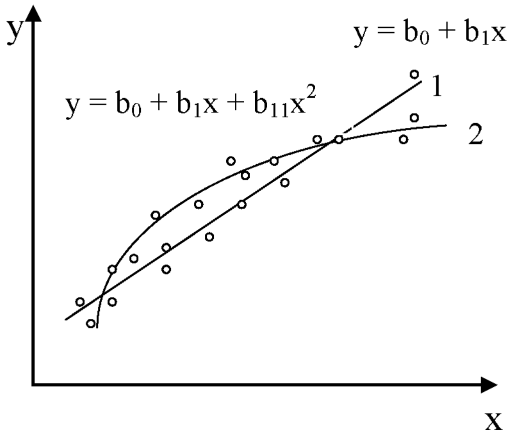

Рис. 1. Апроксимація експериментальних даних лінійною (1) та нелінійною (2) функціями.

• Яка щільність (сила, тіснота, стиснення тощо) зв'язку між **x** та **y** (рис. 2).

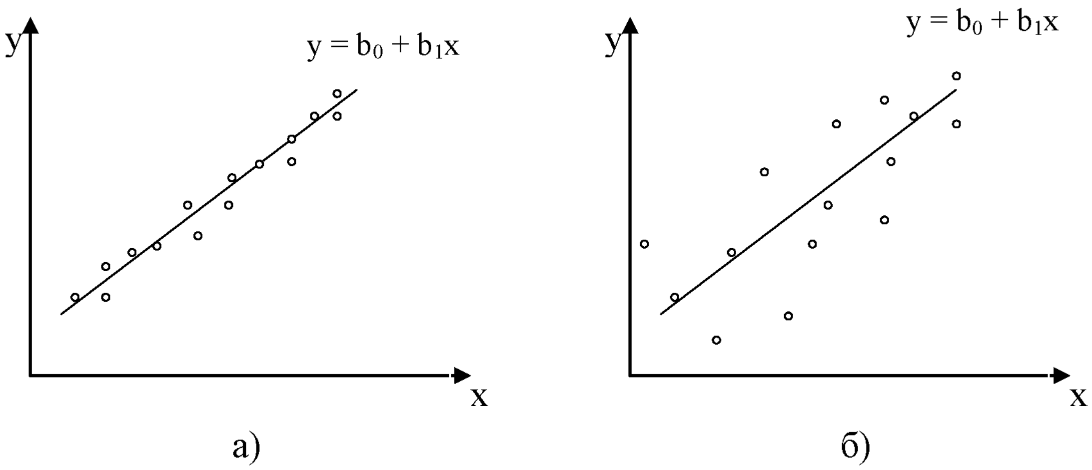

Рис. 2. Сильний (щільний) (а) та слабкий (нещільний) (б) кореляційний лінійний зв'язок *y = f(x)*.

8. Тобто, виникає необхідність виразити щільність (силу) зв'язку певним числом:

а) для лінійного зв'язку це є коефіцієнт кореляції:

• вибірковий коефіцієнт $r = r_{1,2} = r_{x,y} \xrightarrow{\text{ОЦІНКА}}$ ґенеральний коефіцієнт $\rho = \rho_{1,2} = \rho_{x,y}$;

б) для нелінійного зв'язку – кореляційне співвідношення (індекс кореляції):

• вибірковий індекс **R** $\xrightarrow{\text{ОЦІНКА}}$ ґенеральний індекс $\dot{\mathbf{R}}$;

• вибірковий індекс $\mathbf{R}^2$ $\xrightarrow{\text{ОЦІНКА}}$ ґенеральний індекс $\dot{\mathbf{R}}^2$.

1.2. Лінійний зв'язок. Коефіцієнт кореляції

1. Нехай маємо дві ґенеральні сукупности двох випадкових величин **X** та **Y**. Обидві величини мають нормальний закон розподілу Ґавса (табл. 1).

Таблиця 1

Параметри ґенеральних сукупностей величин

Сукупність величин		Параметри
$X:$	$x_1, x_2, x_3, \ldots, x_i, \ldots, x_N$	$N \to \infty, \mu_x, \sigma^2{}_x, \sigma_x, \ldots$
$Y:$	$y_1, y_2, y_3, \ldots, y_i, \ldots, y_N$	$N \to \infty, \mu_y, \sigma^2{}_y, \sigma_y, \ldots$

2. Кількісною оцінкою щільности лінійного зв'язку між величинами ***X*** та **Y** двох ґенеральних сукупностей буде ґенеральний коефіцієнт кореляції:

$$\rho_{x,y} = \rho_{1,2} = \rho = \frac{M_{1/1}}{\sigma_x \cdot \sigma_y} \quad , \tag{2}$$

де $M_{1/1} = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}(x-a_x)(y-a_y)\varphi(x,y)dxdy$ – ґенеральний змішаний централь- (3)
ний момент другого порядку;

$a_x = \int\limits_{-\infty}^{\infty} x\varphi(x)dx$ – математичне сподівання випадкової величини ***X*** … (4)

(ґенеральна середня);

$a_y = \int\limits_{-\infty}^{\infty} y\varphi(y)dy$ – математичне сподівання випадкової величини (5)

Y (ґенеральна середня);

φ(x), φ(y) – щільність ймовірности розподілу випадкової величини ***X*,** **Y** відповідно;

φ(x,y) – щільність ймовірностей двовимірного сумісного розподілу випадкових величин **X**, **Y**;

$\sigma_x = +\sqrt{\sigma_x^2}$, $\sigma_y = +\sqrt{\sigma_y^2}$ – ґенеральні середні квадратичні відхилення (6)

відповідно випадкової величини **X** та **Y**.

Ґенеральні дисперсії визначаються так:

$$\sigma_x^2 = \int\limits_{-\infty}^{\infty}(x-a_x)^2\varphi(x)dx \quad ; \tag{7}$$

$$\sigma_y^2 = \int\limits_{-\infty}^{\infty}(y-a_y)^2\varphi(y)dy \, . \tag{8}$$

3. Коефіцієнт кореляції змінюється у межах:

$$0 \le |\rho| \le 1, \tag{9}$$

$$\text{або} \quad 0\% \le |\rho| \le 100\%; \tag{10}$$

тобто ρ приймає значення: – 1, . . ., 0, . . ., +1; (11)

або ρ: –100%, . . ., 0, . . ., +100%. (12)

4. Числові значення коефіцієнта кореляції визначають вид зв'язку:

• якщо $\rho=0$, то **X**, **Y** є незалежні випадкові величини, тому такі величини некорельовані;

• або, якщо $\rho \to 0$ (у граничному випадку $\rho=0$), то між випадковими величинами **X** та **Y** існує нелінійний зв'язок;

• якщо $|\rho|=1$ (-1, або +1), то зв'язок функціональний;

• якщо $0<|\rho|<1$, то зв'язок стохастичний кореляційний;

• якщо $\rho > 0$, то при $x\uparrow \to y\uparrow$ або при $x\downarrow \to y\downarrow$;

• якщо $\rho < 0$, то при $x\uparrow \to y\downarrow$ або при $x\downarrow \to y\uparrow$.

5. Розрахунок вибіркового коефіцієнта кореляції.

Позначимо вибірковий коефіцієнт кореляції так: $r = r_{1,2} = r_{x,y}$.

Оцінкою генерального коефіцієнта кореляції є вибірковий коефіцієнт кореляції, як і вибіркова сукупність є статистичною оцінкою генеральної сукупности:

вибіркова сукупність $\xrightarrow[\text{ОЦІНКА}]{}$ генеральна сукупність,

а вибіркові характеристики є статистичними оцінками:

$\overline{X} \xrightarrow[\text{ОЦІНКА}]{} \mu_x$ $\qquad$ $\overline{y} \xrightarrow[\text{ОЦІНКА}]{} \mu_y$

$S_x^2 \xrightarrow[\text{ОЦІНКА}]{} \sigma_x^2$ $\qquad$ $S_y^2 \xrightarrow[\text{ОЦІНКА}]{} \sigma_y^2$

$S_x \xrightarrow[\text{ОЦІНКА}]{} S_x$ $\qquad$ $S_y \xrightarrow[\text{ОЦІНКА}]{} S_y$

$m_{1/1} \xrightarrow[\text{ОЦІНКА}]{} M_{1/1}$

. . . . $\qquad$ . . . .

. . . . $\qquad$ . . . .

$$r_{x,y} \xrightarrow[\text{ОЦІНКА}]{} \rho_{x,y}\,.$$

Розраховують вибірковий коефіцієнт кореляції за формулою:

$$r_{x,y} = \frac{m_{1/1}}{S_x \cdot S_y}\ , \tag{13}$$

де $m_{1/1} = \frac{1}{N-1}\sum_{i=1}^{N}\left[(x_i-\overline{x})(y_i-\overline{y})\right] = \frac{1}{N-1}\left[\sum_{i=1}^{N}(x_i\cdot y_i) - \frac{1}{N}\sum_{i=1}^{N}x_i\sum_{i=1}^{N}y_i\right] =$

$= \frac{1}{N-1}\left[\sum_{i=1}^{N}(x_i\cdot y_i) - N(\overline{x})\cdot(\overline{y})\right]$ – вибірковий змішаний центральний (14) момент другого порядку.

Вибіркові середньоквадратичні відхилення:

$$S_x = +\sqrt{S_x^2} = \sqrt{\frac{1}{N-1}\left[\sum_{i=1}^{N}(x_i-\bar{x})^2\right]} = \sqrt{\frac{1}{N-1}\left[\sum_{i=1}^{N}x_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N}x_i\right)^2\right]} = \sqrt{\frac{1}{N-1}\left[\sum_{i=1}^{N}x_i^2 - N\left(\bar{x}\right)^2\right]};$$

(15)

$$S_y = +\sqrt{S_y^2} = \sqrt{\frac{1}{N-1}\left[\sum_{i=1}^{N}(y_i-\bar{y})^2\right]} = \sqrt{\frac{1}{N-1}\left[\sum_{i=1}^{N}y_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N}y_i\right)^2\right]} = \sqrt{\frac{1}{N-1}\left[\sum_{i=1}^{N}y_i^2 - N\left(\bar{y}\right)^2\right]}.$$

(16)

Тоді:

$$r_{x,y} = \frac{\frac{1}{N-1}\sum_{i=1}^{N}\left[(x_i-\bar{x})(y_i-\bar{y})\right]}{\frac{1}{N-1}\sqrt{\sum_{i=1}^{N}(x_i-\bar{x})^2}\cdot\sqrt{\sum_{i=1}^{N}(y_i-\bar{y})^2}} = \frac{\sum_{i=1}^{N}(x_iy_i) - \frac{1}{N}\sum_{i=1}^{N}x_i\sum_{i=1}^{N}y_i}{\sqrt{\left[\sum_{i=1}^{N}x_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N}x_i\right)^2\right]\cdot\left[\sum_{i=1}^{N}y_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N}y_i\right)^2\right]}} =$$

$$= \frac{\sum_{i=1}^{N}(x_iy_i) - N(\bar{x})\cdot(\bar{y})}{\sqrt{\left[\sum_{i=1}^{N}x_i^2 - N\left(\bar{x}\right)^2\right]\cdot\left[\sum_{i=1}^{N}y_i^2 - N\left(\bar{y}\right)^2\right]}}.$$

(17)

1.3. Значущість, довірча ймовірність (надійність) та довірчі інтервали (точність) для коефіцієнта кореляції

1. Вибірковий коефіцієнт кореляції *r* як випадкова величина може приймати різні значення під час повторних досліджень (або спостережень), тобто при «витягуванні» вибірки з ґенеральної сукупности.

Тоді будемо мати різні оцінки ґенерального коефіцієнта кореляції:

2. Навіть, якщо $\rho = 0$ (немає лінійного кореляційного зв'язку), $r \neq 0$. Тоді необхідно перевірити нульову гіпотезу :

$$\left.\begin{array}{l} H_0:\ \rho = 0 \\ \quad\uparrow \text{оцінка} \\ |r| \neq 0 \end{array}\right\} \qquad \text{(немає лінійного кореляційного зв'язку)} \tag{18}$$

Альтернативна гіпотеза:

$$\left.\begin{array}{l} H_1:\ \rho \neq 0 \\ \quad\uparrow \text{оцінка} \\ |r| \neq 0 \end{array}\right\} \qquad \text{(є лінійний кореляційний зв'язок)} \tag{19}$$

3. Перевірка значущости коефіцієнта кореляції за критерієм Стьюдента. При великому обсязі виборки, наприклад N > 100, та $|\rho| \neq 1$ (тобто відсутности функціонального зв'язку), розподіл вибіркового коефіцієнта кореляції **r** підпорядкований нормальному закону розподілу Ґавса (н.з.р.) з параметрами:

$$\left.\begin{array}{l} a_r = \rho; \\ \sigma_r = \dfrac{\sqrt{1-\rho^2}}{\sqrt{N-1}}. \end{array}\right\} \tag{20}$$

Пронормуємо **r**: введемо нову випадкову змінну **z**:

$$z = \frac{r - a_r}{\sigma_r} = \frac{r-\rho}{\sqrt{1-\rho^2}}\sqrt{N-1}. \tag{21}$$

Тоді величина **z** розподілена за нормальним законом розподілу Ґавса з параметрами:

$$\left.\begin{array}{r} \text{математичним сподіванням } a_Z = 0; \\ \text{середнім квадратичним відхиленням } \sigma_Z = 1; \\ \text{дисперсія } \sigma_Z^2 = 1. \end{array}\right\} \tag{22}$$

Так як при N → ∞ (або N → max) $\sigma_r \approx S_r$, то у виразі (21) $\boldsymbol{\sigma_r}$ можна замінити на вибіркове середнє квадратичне відхилення:

$$S_r = \frac{\sqrt{1-r^2}}{\sqrt{N-1}}, \tag{23}$$

тоді отримаємо:

$$z = \frac{r-\rho}{\sqrt{1-r^2}}\sqrt{N-1}. \tag{24}$$

Тоді отримаємо, що **z** підпорядкована розподілу Стьюдента, а перевірка нульової гіпотези H_0: $\rho = 0$ зводиться до розрахунку **статистики**:

$$t_P = \frac{r_P}{\sqrt{1-r_P^2}}\sqrt{N-2}, \tag{25}$$

де $\mathbf{r_P}$ – розрахований коефіцієнт кореляції за вибірковими даними.

Статистика $\mathbf{t_P}$ за правдивости нульової гіпотези $\boldsymbol{H_0}$ має t-розподіл Стьюдента з f=(N–2) ступенями вільностей.

Порівняння $|t_P|$ з теоретичним (табличним) значенням критерію Стьюдента: на рівні значущости $\alpha=1-p$: $t_T\{\alpha;\ f = N-2\}$, де $\alpha=1-p$ – при застосуванні одностороннього критерію та $t_T\{q=1-\alpha/2;\ f = N-2\}$ – при застосуванні двостороннього критерію, приводить до таких висновків:

• якщо: $|t_P| < t_T$, то з ймовірністю правдивости $p = 1-\alpha$ (та ймовірністю ризику $\boldsymbol{\alpha}$ прийняти неправдиву гіпотезу H_0) приймаємо H_0: $\rho = 0$ (немає лінійного зв'язку). Тут оцінка ступеня нелінійности кореляційного зв'язку:

$$\xi_1(t) = \frac{t_T}{|t_P|} > 1, \tag{26}$$

а залишкам ступеня лінійности у нелінійному кореляційному зв'язку можна дати таку оцінку:

$$\xi_2(t) = \frac{|t_P|}{t_T} \le 1; \tag{27}$$

• якщо $|t_P| \ge t_T$, то з ймовірністю правдивости $p = 1-\alpha$ (та ймовірністю ризику $\boldsymbol{\alpha}$ відкинути правдиву гіпотезу H_0) нульова гіпотеза H_0: $\rho = 0$ відкидається, приймається альтернативна гіпотеза H_1: $\rho \neq 0$ (є щільний лінійний зв'язок). Тут оцінка ступеня лінійности кореляційного зв'язку:

$$\xi_2(t) = \frac{|t_P|}{t_T} \ge 1, \tag{28}$$

а залишкам ступеня нелінійности у лінійному кореляційному зв'язку можна дати таку оцінку:

$$\xi_1(t) = \frac{t_T}{|t_P|} < 1. \tag{29}$$

Таким чином, за допомогою коефіцієнта кореляції можна дати (як якісну, так і кількісну) оцінки нелінійного та лінійного зв'язків.

Якщо коефіцієнт кореляції значущий, то для довірчої ймовірности $p=1-\alpha$ довірчий інтервал ґенерального коефіцієнта кореляції становить:

$$p\,\{[\,r-(t_T\,S_r)] \le \rho < [\,r+(t_T S_r)]\} = 1-\alpha\,, \tag{30}$$

де $t_T\{\alpha;\ f = N-2\}$; рівень значущости : $\alpha = (1-p)$ або $t_T\{\ q=1-\alpha/2;\ f = N-2\}$.

4. При малому (обмеженому) обсязі вибірки (N < 60 – 100) і порівняно високої лінійної кореляції розподіл вибіркового коефіцієнта кореляції (**r**) суттєво відрізняється від нормального закону розподілу Ґавса (рис. 3).

Тоді, використовуючи перетворення Фішера, введемо нову випадкову величину:

$$Z = \frac{1}{2}\ln\frac{1+\rho}{1-\rho}, \tag{31}$$

де $\boldsymbol{\rho}$ – ґенеральний коефіцієнт кореляції.

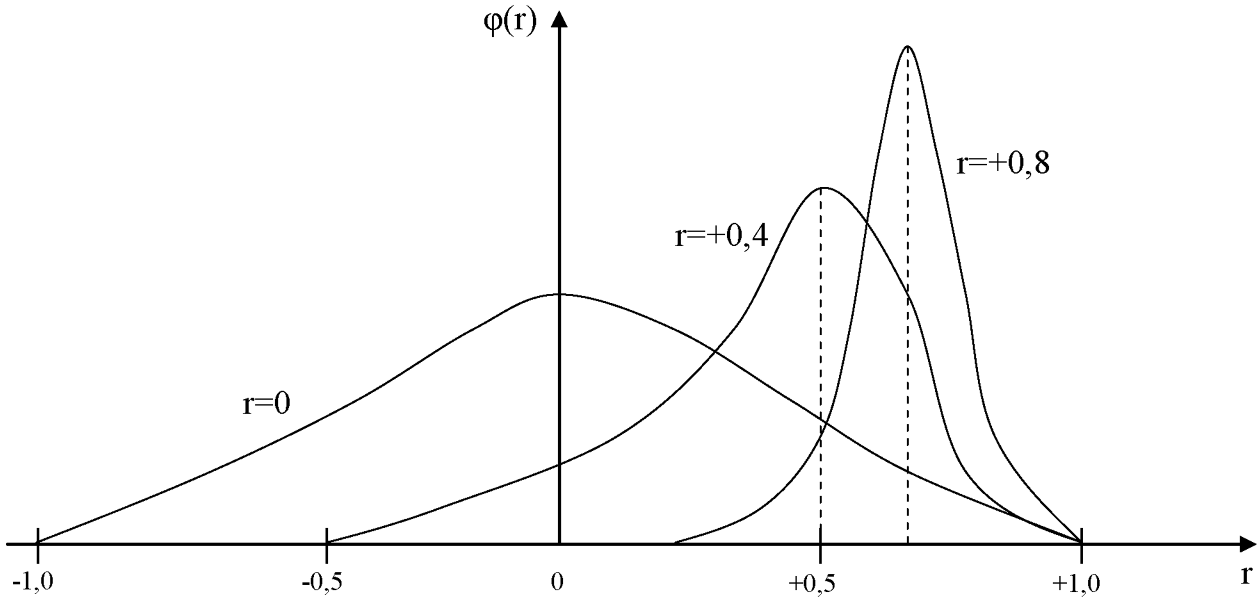

Рис. 3. Розподіл вибіркового коефіцієнта кореляції при малому обсязі виборки.

Ця випадкова величина Z, як показав Фішер, розподілена за нормальним законом розподілу Ґавса (рис. 4) з параметрами:

• математичним сподіванням:

$$a_Z = \frac{1}{2}\ln\frac{1+\rho}{1-\rho} + \frac{\rho}{2(N-1)}; \tag{32}$$

• середнім квадратичним відхиленням:

$$\sigma_Z = \frac{1}{\sqrt{N-3}}. \tag{33}$$

Так як [$\frac{\rho}{2(N-2)}$ <<σ_Z], то в (32) цим доданком можна знехтувати, тоді:

$$a_Z \approx \frac{1}{2}\ln\frac{1+\rho}{1-\rho}. \tag{34}$$

Перевірка нульової гіпотези H_0: $\rho = 0$ зводиться до розрахунку статистики:

$$Z_P = \frac{1}{2}\ln\frac{1+r_P}{1-r_P}, \quad r \neq 0 \tag{35}$$

яка добре апроксимується нормальним законом розподілу Ґавса з математичним сподіванням (32) та середнім квадратичним відхиленням (33), і порівняння $|Z_P|$ з добутком [(Z_T {q = $1 - \frac{\alpha}{2}$ } $\cdot\, \sigma_Z$)], де Z_T – квантиль нормованого нормального розподілу для ймовірності p = $1 - \alpha$, при цьому:

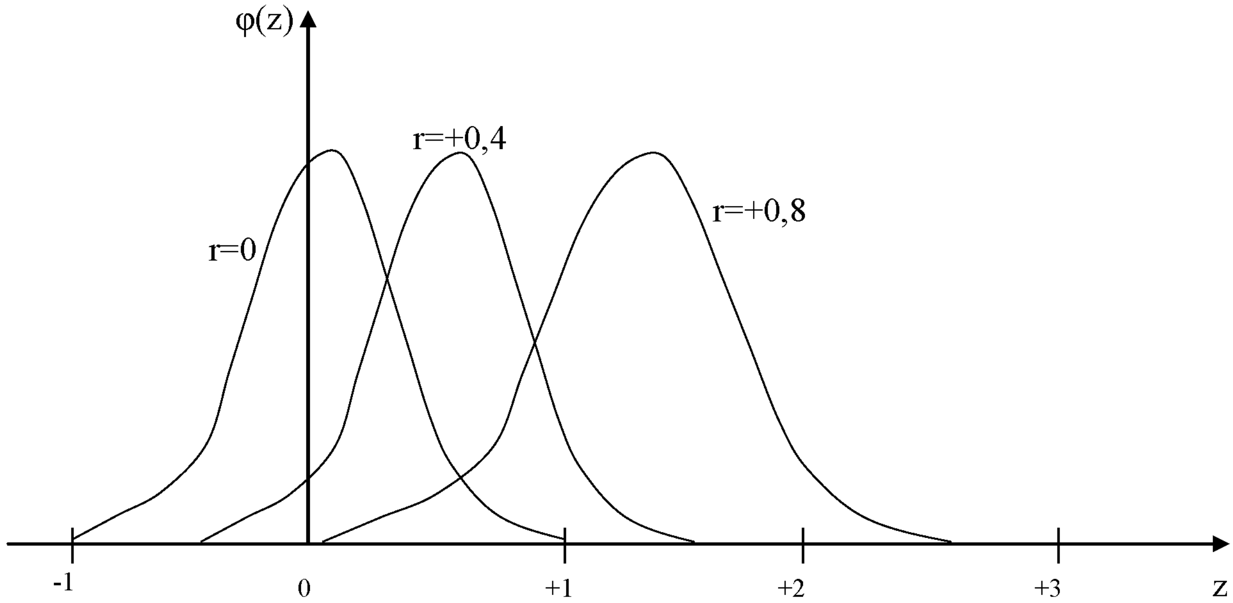

Рис. 4. Нормальний розподіл випадкової величини Z.

• **якщо** $|Z_P|$ **<** **(** Z_T **{ q =** ***1*** **–** $\frac{\alpha}{2}$ **}** $\cdot \sigma_Z$ **),** то нульову гіпотезу H_0: $\rho = 0$ (немає лінійного зв'язку) приймаємо з ймовірністю правдивости р = $1 - \alpha$ та ймовірністю ризику ***α*** прийняти неправдиву гіпотезу $\boldsymbol{H_0}$, тут оцінка ступеня нелінійности кореляційного зв'язку:

$$\xi_1(Z) = \frac{[Z_T \cdot \sigma_Z]}{|Z_P|} > 1, \tag{36}$$

а залишкам ступеня лінійности у нелінійному кореляційному зв'язку можна дати таку оцінку:

$$\xi_2(Z) = \frac{|Z_P|}{[Z_T \cdot \sigma_Z]} \le 1; \tag{37}$$

• **якщо** $|Z_P|$ **≥** **[(** Z_T **{ q =** ***1*** **–** $\frac{\alpha}{2}$ **}** $\cdot \sigma_Z$ **)]**, то нульова гіпотеза $\boldsymbol{H_0}$ відкидається з ймовірністю правдивости відкидання $\boldsymbol{H_0}$ р = $1 - \alpha$ та ймовірністю ризику ***α*** відкинути правдиву гіпотезу $\boldsymbol{H_0}$, приймається альтернативна гіпотеза H_1: $\rho \neq 0$ (існує щільний лінійний зв'язок – коефіцієнт кореляції значущий), тут оцінка ступеня лінійности кореляційного зв'язку:

$$\xi_2(Z) = \frac{|Z_P|}{[Z_T \cdot \sigma_Z]} \ge 1, \tag{38}$$

а залишкам ступеня нелінійности у лінійному кореляційному зв'язку можна дати таку оцінку:

$$\xi_1(Z) = \frac{[Z_T \cdot \sigma_Z]}{|Z_P|} < 1. \tag{39}$$

При цьому довірчі інтервали та довірча ймовірність для математичного сподівання a_z:

$$p\,[\,(Z_1) < a_z < (Z_2)\,] = 1 - \alpha, \tag{40}$$

$$\text{де } Z_1 = Z - (Z_T \cdot \sigma_Z); \tag{41}$$

$$Z_2 = Z + (Z_T \cdot \sigma_Z); \tag{42}$$

Довірчі інтервали та довірча ймовірність для генерального коефіцієнта кореляції ρ:

$$p\ [\ r_1 < \rho < r_2\ ] = 1 - \alpha\ , \qquad (43)$$

$$\left.\begin{aligned} r_1 &= thz_1 \\ r_2 &= thz_2 \end{aligned}\right\} \qquad (44)$$

сіну

де

$$th(z) = \frac{\sinh(z)}{\cosh(z)} = \frac{\dfrac{\exp(z)-\exp(-z)}{2}}{\dfrac{\exp(z)+\exp(-z)}{2}} = \frac{\exp(z)-\exp(-z)}{\exp(z)+\exp(-z)} = \frac{\dfrac{\exp(z)}{\exp(-z)}-1}{\dfrac{\exp(z)}{\exp(-z)}+1} = \frac{\exp(2z)-1}{\exp(2z)+1}, \qquad (45)$$

звідки

$$r_1 = thz_1 = \frac{\exp(2z_1)-1}{\exp(2z_1)+1}; \qquad (46)$$

$$r_2 = thz_2 = \frac{\exp(2z_2)-1}{\exp(2z_2)+1}. \qquad (47)$$

де th(z) = tgh(z) = tanh(z) = Th(z) – гіперболічний тангенс [10]:

th(z) = Sh(z)/Ch(z),

тут: Sh(z) = [exp(z)-exp(-z)]/2=sinh(z) гіперболічний сінус;

Ch(z) = [exp(z)+exp(-z)]/2=cosh(z) гіперболічний косинус.

5. Перевірку значущости коефіцієнта кореляції проводять також за його нижньою межею довірчої ділянки для абсолютного значення коефіцієнта кореляції:

$r_{KP}\ \{\ \alpha = 1 - p;\ f = N - 2\}$ при застосуванні одностороннього критерію; (48)

$r_{KP}\ \{\ q = 1 - \dfrac{\alpha}{2}\ ;\ f = N - 2\}$ при застосуванні двостороннього критерію.

Вибірковий коефіцієнт кореляції **r** зв'язаний з випадковою величиною **t**, яка підпорядкована розподілу Стьюдента з $f = N - 2$.

в) За перетвореннями Фішера змінна

$$t = |r|\sqrt{\frac{N-2}{1-r^2}} \qquad (49)$$

дійсно підпорядковується розподілу Стьюдента. Тому розраховується статистика:

$$t_P = |r|\sqrt{\frac{N-2}{1-r^2}} \qquad (50)$$

і порівнюється з t_T {α=1–p, $f = N - 2$} або t_T {q=1–$\dfrac{\alpha}{2}$, $f = N - 2$}, виходячи із співвідношення:

$$\frac{t_T}{\sqrt{N-2}} = \frac{|r|}{\sqrt{1-r^2}}, \qquad (51)$$

звідки визначається межа, нижче якої **r** вже не відрізняється від ρ = 0:

$$r = \frac{t_T}{\sqrt{f + t_T^2}}. \qquad (52)$$

г) Критичне значення $r_{кр}$ {α=1–p, f = *N* – 2} або $r_{кр}$ {q=1–$\frac{\alpha}{2}$, f = *N* – 2} може бути визначено через квантиль t_T {α=1–p, f = *N* – 2} (під час застосування одностороннього критерію **t**), або через квантиль t_T {q=1–$\frac{\alpha}{2}$, f = *N* – 2} (під час застосування двостороннього критерію) порядка p=1–α розподілу Стьюдента за виразом:

$$r_{кр} = r\{\alpha = 1 - p; f = N - 2\} = \frac{t_T\{\alpha = 1 - p; f = N - 2\}}{\sqrt{f + t_T\{\alpha = 1 - p; f = N - 2\}}}, \qquad (53)$$

або $$r_{кр} = r\{q = 1 - \frac{\alpha}{2}; f = N - 2\} = \frac{t_T\{q = 1 - \frac{\alpha}{2}; f = N - 2\}}{\sqrt{f + t_T\{q = 1 - \frac{\alpha}{2}; f = N - 2\}}}.$$

(54)

1) **При застосуванні одностороннього критерію** на рівні значущости α=1–p ***H₀*** приймається, якщо $|r|$ < $r_{КР}$ {α=1–p, f = *N* – 2} (немає значущого лінійного зв’язку), тут оцінка ступеня нелінійности кореляційного зв’язку:

$$\xi_1(r) = \frac{r_{КР}}{|r|} > 1, \qquad (55)$$

а залишкам ступеня лінійности у нелінійному кореляційному зв’язку можна дати таку оцінку:

$$\xi_2(r) = \frac{|r|}{r_{КР}} \le 1; \qquad (56)$$

і ***H₀*** відкидається, якщо $|r|$ ≥ $r_{КР}$ {α=1–p, f = *N* – 2} (є значущий лінійний зв’язок), тут оцінка ступеня лінійности кореляційного зв’язку:

$$\xi_2(r) = \frac{|r|}{r_{КР}} \ge 1, \qquad (57)$$

а залишкам ступеня нелінійности у лінійному кореляційному зв’язку можна дати таку оцінку:

$$\xi_1(r) = \frac{r_{КР}}{|r|} < 1. \qquad (58)$$

2) При застосуванні двостороннього критерію на рівні значущости α=1–p нульову гіпотезу ***H₀*** приймають, якщо $|r|$ < $r_{КР}$ {q=1–$\frac{\alpha}{2}$, f = *N* – 2} і відкидають, якщо $|r|$ > $r_{КР}$ {q=1–$\frac{\alpha}{2}$, f = *N* – 2} з оцінками ступеня лінійности і нелінійности [(55) (58)].

1.4. Числовий приклад №1: лінійна кореляція між двома величинами

1. За літературними даними [1] шукали залежність приросту ентальпії $\Delta H_T = H_T - H_0$ (де H_0 – ентальпія при T = 0 K; H_T – ентальпія при T > 0 K) металів від температури (T), радіусу атомів (r_A) та порядкового номера (Z) в Періодичній таблиці первнів. У табл. 2 зведені вихідні значення зміни ентальпії за температури 298,15 K; 400 K; 900 K, а також порядковий номер (Z) та радіус атома (r_A) для 36 металів.

2. Висували нульові гіпотези рівности нулю ґенеральних коефіцієнтів кореляцій:

$$H_0: \rho\,[\Delta H_T = f(Z)] = 0; \tag{59}$$

$$\left.\begin{array}{l} H_0: \rho\,[\Delta H_T = f(r_A)] = 0. \end{array}\right\}$$

Альтернативні гіпотези:

$$\left.\begin{array}{l} H_1: \rho\,[\Delta H_T = f(Z)] \neq 0; \\ H_1: \rho\,[\Delta H_T = f(r_A)] \neq 0. \end{array}\right\} \tag{60}$$

У табл. 3 зведені результати перевірки нульової гіпотези $\mathbf{H_0}$ рівности нулю ґенерального коефіцієнта кореляції за $r_{кр}$, t_T, $(z_T \cdot \sigma_z)$ за рівнем значущости α = 0,05 та α = 0,01 [2].

а) Як видно з табл. 3, для залежности приросту ентальпії від порядкового номера первня **Z**, $\Delta H_T = f(Z)$ ступені лінійности становлять: $\xi_2(r)>1$; $\xi_2(t)>1$; $\xi_2(z)>1$, тобто нульова гіпотеза $\mathbf{H_0}$ відкидається – між $\mathbf{\Delta H_T}$ та **Z** є щільний лінійний кореляційний зв'язок з ймовірностями p = 1 – α = 0,95 та p = 0,99.

б) Як видно з табл. 3, для залежности приросту ентальпії від радіусу атома $\mathbf{r_A}$ первня $\Delta H_T = f(r_A)$ ступені нелінійности становлять: $\xi_1(r)>1$; $\xi_1(t)>1$; $\xi_1(z)>1$, тобто нульова гіпотеза $\mathbf{H_0}$ приймається – між $\mathbf{\Delta H_T}$ та $\mathbf{r_A}$ є нелінійний кореляційний зв'язок з ймовірностями p = 1 – α = 0,95 та p = 0,99.

1.5. Порівняння двох ґенеральних коефіцієнтів кореляції

1. Іноді виникає необхідність порівняти два вибіркових коефіцієнта кореляції з метою перевірки нульової гіпотези про відсутність татистичної різниці між двома ґенеральними коефіцієнтами кореляції $\boldsymbol{\rho_1}$ і $\boldsymbol{\rho_2}$, які оцінюються за двома відповідними вибірковими коефіцієнтами кореляції $\mathbf{r_1}$ і $\mathbf{r_2}$:

$$\left.\begin{array}{l} H_0\!: \rho_1 \;=\; \rho_2 \\ \quad\;\;\uparrow \qquad \uparrow \;\; \text{оцінка} \\ \quad\; r_1 \;\neq\; r_2 \end{array}\right\} \tag{61}$$

2. Розраховують статистики:

$$Z_1 = \frac{1}{2}\ln\frac{1+r_1}{1-r_1}; \tag{62}$$

$$Z_2 = \frac{1}{2}\ln\frac{1+r_2}{1-r_2}; \tag{63}$$

та сумісну статистику:

$$Z_P = \left|\frac{Z_1 - Z_2}{\sqrt{\dfrac{1}{N_1-3}+\dfrac{1}{N_2-3}}}\right|. \tag{64}$$

Таблиця 2

Залежність зміни ентальпії металів від температури, радіусу атомів та порядкового номера в Періодичній таблиці первнів

Метал	Порядковий номер, Z	Радіус атома r_A, нм	ΔH, кДж/кг		
			298,15K	400 K	900 K
Li	3	0,157	667,5	1049,1	–
Na	11	0,191	281	–	–
K	19	0,235	181,3	–	–
Rb	37	0,25	87,62	–	–
Cs	55	0,272	58,02	–	–
Cu	29	0,128	–	39,5	250
Ag	47	0,144	–	23,8	150
Au	79	0,146	–	13,3	81,7
Be	4	0,112	216,4	422,1	1727,3
Mg	12	0,16	205,6	312,5	906,1
Ca	20	0,197	143	209,8	623,1
Sr	38	0,215	74,98	106,4	293,3
Ba	56	0,221	50,31	74,29	224,8
Zn	30	0,138	–	40	–
Cd	48	0,155	–	24	–
Al	13	0,143	–	263,8	799,7
Sc	21	0,162	–	174,6	486,9
Y	39	0,178	–	97,81	260,95
Ti	22	0,147	–	155,6	456,5
Zr	40	0,16	68,2	101,2	285,3
Hf	72	0,167	–	47,57	126,6
Sn	50	0,162	–	50	–
Pb	82	0,175	33,2	46,4	–
V	23	0,134	–	140,1	408,8
Nb	41	0,146	–	83,61	226,9
Ta	73	0,149	–	45,82	121,2
Bi	83	0,1545	–	13	–
Cr	24	0,13	–	125,6	389,8
Mo	42	0,139	–	73,98	212,8
W	74	0,141	–	40,71	111,2
Mn	25	0,135	–	52,5	348
α-Fe	26	0,126	80,2	128	430
γ-Fe	26	0,126	200	254	528
Co	27	0,125	–	45,4	311
Ni	28	0,124	–	47,4	317
Pt	78	0,137	–	13,8	84,7

Таблиця 3

Результати перевірки нульової гіпотези рівности нулю ґенерального коефіцієнта кореляції зв’язків $\Delta H_T \sim Z$ та $\Delta H_T \sim r_A$ за його критичним значенням ($r_{кр.}$), критерієм Стьюдента (t_T) та перетворенням Фішера (z_T)

Зв’язки між величинами		Рівні значущости	ΔH_T, кДж/кг		
			298,15 K	400 K	900 K
Кореляційний зв’язок:			Розрахунковий коефіцієнт кореляції (r_p)		
$\Delta H_T \sim Z$			−0,6808	0,5822	−0,7226
$\Delta H_T \sim r_A$			−0,3100	−0,0141	−0,2083
N			14	38	26
f = N−2			12	36	24
			Критичний коефіцієнт кореляції		
$r_{кр.}$ $\{q = 1 - \frac{\alpha}{2}$; f =N−2}		α = 0,05	0,5324	0,3202	0,3882
		α = 0,01	0,6614	0,4128	0,4958
$\Delta H_T \sim Z$	ξ_1 (r)	α = 0,05	0,782	0,550	0,537
	ξ_2 (r)		1,279	1,818	1,861
	ξ_1 (r)	α = 0,01	0,972	0,709	0,682
	ξ_2 (r)		1,029	1,410	1,466
$\Delta H_T \sim r_A$	ξ_1 (r)	α = 0,05	1,717	22,71	1,864
	ξ_2 (r)		0,582	0,044	0,537
	ξ_1 (r)	α = 0,01	2,134	29,28	2,380
	ξ_2 (r)		0,469	0,034	0,420
Кореляційний зв’язок:			Статистика Стьюдента (t_p)		
$\Delta H_T \sim Z$			−3,2197	–4,2964	−5,1211
$\Delta H_T \sim r_A$			−1,1295	−0,0846	−1,0433
t_T $\{q = 1 - \frac{\alpha}{2}$; f = N−2}		α = 0,05	2,179	2,028	2,064
		α = 0,01	3,055	2,719	2,797
$\Delta H_T \sim Z$	ξ_1 (t)	α = 0,05	0,677	0,472	0,403
	ξ_2 (t)		1,478	2,119	2,481
	ξ_1 (t)	α = 0,01	0,949	0,633	0,546
	ξ_2 (t)		1,054	1,580	1,831
$\Delta H_T \sim r_A$	ξ_1 (t)	α = 0,05	1,929	23,97	1,978
	ξ_2 (t)		0,518	0,042	0,505
	ξ_1 (t)	α = 0,01	2,705	32,14	2,681
	ξ_2 (t)		0,370	0,031	0,373
Кореляційний зв’язок:			Статистика перетворення Фішера (z_p)		
$\Delta H_T \sim Z$			−0,8306	0,6658	−0,9131
$\Delta H_T \sim r_A$			−0,3205	−0,0141	−0,2114
z_T $\{q = 1 - \frac{\alpha}{2}\}$		α = 0,05	1,96	1,96	1,96
		α = 0,01	2,58	2,58	2,58
σ_Z			0,3015	0,1690	0,2085
$(z_T \cdot \sigma_Z)$		α = 0,05	0,5910	0,3313	0,4087
		α = 0,01	0,7779	0,4361	0,5380
$\Delta H_T \sim Z$	ξ_1 (Z)	α = 0,05	0,712	0,498	0,448
	ξ_2 (Z)		1,405	2,010	2,234
	ξ_1 (Z)	α = 0,01	0,937	0,655	0,589
	ξ_2 (Z)		1,068	1,527	1,697
$\Delta H_T \sim r_A$	ξ_1 (Z)	α = 0,05	1,844	23,50	1,933
	ξ_2 (Z)		0,542	0,043	0,517
	ξ_1 (Z)	α = 0,01	2,427	30,93	2,545
	ξ_2 (Z)		0,412	0,032	0,393

3. Із відповідної таблиці знаходять квантиль нормованого нормального розподілу Ґавса для p = 1 – α :

$$Z_T\ \{\ \mathrm{q} = 1 - \frac{\alpha}{2}\ \}. \tag{65}$$

4. Якщо $|Z_P| \le Z_T\ \{\ \mathrm{q} = 1 - \frac{\alpha}{2}\ \}$, **то нульова гіпотеза** $\boldsymbol{H_0}$ приймаємо. Тобто з рівнем значущости **α** (для ймовірности p = 1 – α) стверджуємо, що між $\boldsymbol{\rho_1}$ і $\boldsymbol{\rho_2}$ немає статистичної різниці. Тобто можна вважати, що вибірки взяті із загальної сукупности.

Ступінь рівности двох коефіцієнтів кореляції:

$$\xi_1(z) = \frac{Z_T}{|Z_P|} \ge 1 \tag{66}$$

та із залишками ступеня нерівности:

$$\xi_2(z) = \frac{|Z_P|}{Z_T} < 1. \tag{67}$$

5. Якщо $|Z_P| > Z_T$**, то стверджуємо**, що між $\boldsymbol{\rho_1}$ і $\boldsymbol{\rho_2}$ є статистична різниця зі ступенем нерівности:

$$\xi_2(z) = \frac{|Z_P|}{Z_T} > 1 \tag{68}$$

та із залишками ступеня рівности:

$$\xi_1(z) = \frac{Z_T}{|Z_P|} \le 1. \tag{69}$$

2. Лінійна множинна кореляція та реґресія

2.1. Теоретична частина

1. У випадку k-змінних рівняння множинної реґресії у натуральній скалі має вигляд:

$$\bar{y}_1(2,3,...,i,...,k) = b_1 + b_2x_2 + b_3x_3 + ... + b_ix_i + .. + b_kx_k\ , \tag{70}$$

де $\bar{\mathrm{y}}_1(2,3,...,\mathrm{i},...,\mathrm{k})$ – умовне середнє значення залежної величини $\mathbf{y_1}$, яке відповідає певним значенням незалежних величин $x_2, x_3, \ldots, x_i, \ldots, x_k$.

2. Проведемо процедуру нормування: перейдемо до нової випадкової нормованої змінної $\mathbf{t_i}$:

$$\left.\begin{aligned} \mathrm{t_i} &= \frac{\mathrm{y_j} - \bar{\mathrm{y}}_\mathrm{i}}{\mathrm{S_{yi}}}; \\ \mathrm{t_j} &= \frac{\mathrm{x_j} - \bar{\mathrm{x}}_\mathrm{i}}{\mathrm{S_{xi}}}, \end{aligned}\right\}, \tag{71}$$

тоді всі величини $\mathbf{y_i}$ та всі їх залежности знайдуть вираження у стандартній скалі.

3. Таким чином, рівняння реґресії (70) в нормованому вигляді

набуде форми:

$$\bar{t}_1(2,3,...,i,...,k) = \beta_2 t_2 + \beta_3 t_3 + ... + \beta_i t_i + .. + \beta_k t_k \,, \qquad (72)$$

де $\bar{t}_1(2,3,...,i,...,k)$ – умовне середнє значення нормованої (стандартної) залежної величини $\mathbf{t_1}$, яке відповідає певним значенням нормованих незалежних величин t_2, t_3,…, t_i,…, $t_к$ де:

t_2, t_3,…, t_i,…, $t_к$ – значення нормованих (стандартних) незалежних величин y_2,y_3,…,y_i,…,$y_к$;

β_2, β_3,…, β_i,…, $\beta_к$ – нормовані (стандартні) коефіцієнти множинної регресії за рівнянням (72).

4. Нормовані (стандартні) коефіцієнти множинної регресії визначимо за системою лінійних рівнянь:

$$\left.\begin{array}{l}
r_{1,2} = \beta_2 r_{22} + \beta_3 r_{32} + \beta_4 r_{42} + ... + \beta_i r_{i2} + ... + \beta_k r_{k2} \\
r_{1,3} = \beta_2 r_{23} + \beta_3 r_{33} + \beta_4 r_{43} + ... + \beta_i r_{i3} + ... + \beta_k r_{k3} \\
r_{1,4} = \beta_2 r_{24} + \beta_3 r_{34} + \beta_4 r_{44} + ... + \beta_i r_{i4} + ... + \beta_k r_{k4} \\
\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
r_{1,i} = \beta_2 r_{2i} + \beta_3 r_{3i} + \beta_4 r_{4i} + ... + \beta_i r_{ii} + ... + \beta_k r_{ki} \\
\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
r_{1,k} = \beta_2 r_{2k} + \beta_3 r_{3k} + \beta_4 r_{4k} + ... + \beta_i r_{ik} + ... + \beta_k r_{kk},
\end{array}\right\} , \qquad (73)$$

де r_{12}, r_{13}, r_{14},…, r_{1i},… r_{1k},…, r_{22}, r_{33}, r_{ii}, r_{32}, r_{42}, r_{23}, r_{43},…, r_{ik},…, r_{kk} – коефіцієнти парної лінійної кореляції між змінними y_1~x_2; y_1~x_3; x_2~x_3; x_2~x_4;…; x_i~x_i; x_i~$x_к$;…; $x_к$~$x_к$.

5. Щільність зв'язку змінної y_1 із сукупністю змінних x_2, x_3,…, x_i,…, $x_к$ у випадку лінійної множинної кореляції визначається **за коефіцієнтом множинної кореляції**:

$$r_1(2,3,...,i,...,k) = \sqrt{\beta_2 r_{12} + \beta_3 r_{13} + ... + \beta_i r_{1i} + ... + \beta_k r_{1k}} \,. \qquad (74)$$

6. Коефіцієнт множинної кореляції після корекції (врахування числа коефіцієнтів рівняння (70) – числа параметрів рівняння регресії в натуральній скалі дорівнює:

$$\bar{r}_1(2,3,...,i,...,k) = \sqrt{1 - \left[1 - r_1^2(2,3,...,i,...,k)\right] \cdot \left(\frac{N-1}{N-k}\right)} \,, \qquad (75)$$

де N – число спостережень;

k – число параметрів (коефіцієнтів) математичної моделі (70).

7. Розрахунок коефіцієнтів математичної моделі (70) в натуральній скалі:

$$b_i = \beta_1 \frac{S_{y1}}{S_i}, \text{ де } i = 2, 3, \dots, i, \dots, k; \qquad (76)$$

$$b_1 = \bar{y}_1 - \left[b_2 \bar{x}_2 + b_3 \bar{x}_3 + ... + b_i \bar{x}_i + ... + b_k \bar{x}_k\right], \qquad (77)$$

де $\bar{y}_1, \bar{x}_2, \bar{x}_3, ...\bar{x}_i, ...\bar{x}_k$ – середнє значення відповідних величин.

8. Оцінка середньої квадратичної помилки розрахунку величини $\bar{y}_1(2,3,...,i,..,k)$ в рівнянні моделі (70) дорівнює:

$$\delta_{y1} = \sqrt{S^2(y_1)} = S(y_1)\sqrt{\left[1-\bar{r}_1^2(2,3,...,i,...,k)\right]\cdot\left(\frac{N-1}{N-2}\right)}. \quad (78)$$

2.2. Числовий приклад №2: лінійна множинна кореляція та реґресія

В основі цього розділу покладений числовий приклад, взятий з [3, 6].

Середнє значення та середні квадратичні відхилення границі міцности зразків алюмінієвого стопу АМг6 (N = 26 топів) та концентрації основних інгредієнтів Mg, Fe, Si у цьому стопі приведені в табл. 4. Визначимо коефіцієнти моделі (70) у натуральній скалі (емпіричної регресії) для границі міцности стопу (**y_1**) як функції від концентрації основних інгредієнтів: Mg (**x_2**), Fe (**x_3**), Si (**x_4**) та коефіцієнт множинної кореляції.

Таблиця 4

Статистичні характеристики границі міцности та концентрацій Mg, Fe, Si в алюмінієвому стопі АМг6

Випадкова величина	Натуральне позначення	Розмірність	Кодоване позначення	Середня $\bar{y}_1, \bar{x}_i$	Середнє квадратичне відхилення S_i
границя міцности	$\sigma_в$	МПа	y_1	359,7079	5,8860
концентрація Mg	C (Mg)	%	x_2	6,34	0,1136
концентрація Fe	C (Fe)	%	x_3	0,30	0,0200
концентрація Si	C (Si)	%	x_4	0,22	0,0500

1. Розрахунок коефіцієнтів парних кореляцій r_{ij} між границею міцности стопу **y_1** та концентраціями первнів **x_2, x_3, x_4** (Mg, Fe, Si) зведені у табл. 5 (у загальному вигляді) та у табл. 6 (у числових значеннях).

2. За формулою (70) рівняння емпіричної реґресії в натуральній скалі для даного прикладу має вигляд:

$$\bar{y}_1(2,3,4) = b_1 + b_2x_2 + b_3x_3 + b_4x_4 . \quad (79)$$

3. За формулою (72) у нормованій скалі рівняння (79) набуває вигляду:

$$\bar{t}_1(2,3,4) = \beta_2t_2 + \beta_3t_3 + \beta_4t_4 . \quad (80)$$

4. Знайдемо коефіцієнти β_2, β_3, β_4, розв’язуючи систему лінійних рівнянь:

$$\left.\begin{aligned} r_{1,2} &= \beta_2r_{22} + \beta_3r_{32} + \beta_4r_{42}; \\ r_{1,3} &= \beta_2r_{23} + \beta_3r_{33} + \beta_4r_{43}; \\ r_{1,4} &= \beta_2r_{24} + \beta_3r_{34} + \beta_4r_{44}. \end{aligned}\right\} \quad (81)$$

Таблиця 5

Матриця вибіркових коефіцієнтів парних кореляцій r_{ij}

Випадкова величина	Коефіцієнт парної кореляції			
	y_1 [$\sigma_в$]	x_2 [C (Mg)]	x_3 [C (Fe)]	x_4 [C (Si)]
y_1 [$\sigma_в$]	r_{11}	r_{12}	r_{13}	r_{14}
x_2 [C (Mg)]	r_{21}	r_{22}	r_{23}	r_{24}
x_3 [C (Fe)]	r_{31}	r_{32}	r_{33}	r_{34}
x_4 [C (Si)]	r_{41}	r_{42}	r_{43}	r_{44}

Таблиця 6

Значення вибіркових коефіцієнтів парних кореляцій r_{ij}між границею міцности та концентраціями Mg, Fe, Si

Випадкова величина	Коефіцієнт парної кореляції			
	y_1 [$\sigma_в$]	x_2 [C (Mg)]	x_3 [C (Fe)]	x_4 [C (Si)]
y_1 [$\sigma_в$]	1	0,5352	–0,4273	–0,2659
x_2 [C (Mg)]	0,5352	1	–0,4286	–0,6458
x_3 [C (Fe)]	–0,4273	–0,4286	1	0,6154
x_4 [C (Si)]	–0,2659	–0,6458	0,6154	1

5. Для системи лінійних рівнянь (81) знайдемо визначники:

$$\Delta_0 = \begin{vmatrix} \mathbf{r_{22}} & \mathbf{r_{32}} & \mathbf{r_{42}} \\ \mathbf{r_{23}} & \mathbf{r_{33}} & \mathbf{r_{43}} \\ \mathbf{r_{24}} & \mathbf{r_{34}} & \mathbf{r_{44}} \end{vmatrix} = \begin{vmatrix} \mathbf{1} & \mathbf{-0,4286} & \mathbf{-0,6458} \\ \mathbf{-0,4286} & \mathbf{1} & \mathbf{0,6154} \\ \mathbf{-0,6458} & \mathbf{0,6154} & \mathbf{1} \end{vmatrix} = \mathbf{+0,3611}; \qquad (82)$$

$$\Delta_2 = \begin{vmatrix} \mathbf{r_{12}} & \mathbf{r_{32}} & \mathbf{r_{42}} \\ \mathbf{r_{13}} & \mathbf{r_{33}} & \mathbf{r_{43}} \\ \mathbf{r_{14}} & \mathbf{r_{34}} & \mathbf{r_{44}} \end{vmatrix} = \begin{vmatrix} \mathbf{0,5352} & \mathbf{-0,4286} & \mathbf{-0,6458} \\ \mathbf{-0,4273} & \mathbf{1} & \mathbf{0,6154} \\ \mathbf{-0,2659} & \mathbf{0,6154} & \mathbf{1} \end{vmatrix} = \mathbf{+0,2176}; \qquad (83)$$

$$\Delta_3 = \begin{vmatrix} \mathbf{r_{22}} & \mathbf{r_{12}} & \mathbf{r_{42}} \\ \mathbf{r_{23}} & \mathbf{r_{13}} & \mathbf{r_{43}} \\ \mathbf{r_{24}} & \mathbf{r_{14}} & \mathbf{r_{44}} \end{vmatrix} = \begin{vmatrix} \mathbf{1} & \mathbf{0,5352} & \mathbf{-0,6458} \\ \mathbf{-0,4286} & \mathbf{-0,4273} & \mathbf{0,6154} \\ \mathbf{-0,6458} & \mathbf{-0,2659} & \mathbf{1} \end{vmatrix} = \mathbf{-0,1424}; \qquad (84)$$

$$\Delta_4 = \begin{vmatrix} r_{22} & r_{32} & r_{12} \\ r_{23} & r_{33} & r_{13} \\ r_{24} & r_{34} & r_{14} \end{vmatrix} = \begin{vmatrix} 1 & -0{,}4286 & 0{,}5352 \\ -0{,}4286 & 1 & -0{,}4273 \\ -0{,}6458 & 0{,}6154 & -0{,}2659 \end{vmatrix} = +0{,}1320. \quad (85)$$

6. Тоді коефіцієнти рівняння (81) будуть дорівнювати:

$$\begin{cases} \beta_2 = \dfrac{\Delta_2}{\Delta_0} = \dfrac{+0{,}2176}{+0{,}3611} = +0{,}6026; & (86) \\[2ex] \beta_3 = \dfrac{\Delta_3}{\Delta_0} = \dfrac{-0{,}1424}{+0{,}3611} = -0{,}3944; & (87) \\[2ex] \beta_4 = \dfrac{\Delta_4}{\Delta_0} = \dfrac{+0{,}1320}{+0{,}3611} = +0{,}3655. & (88) \end{cases}$$

7. Рівняння (80) у нормованій скалі набуває вигляду:

$$\bar{t}_1(2,3,4) = 0{,}6026t_2 - 0{,}3944t_3 + 0{,}3655t_4. \quad (89)$$

8. Розрахунок коефіцієнтів моделі (79) у натуральній скалі за формулами (76), (77), (90)–(92) та табл. 4 привів до таких результатів:

$$b_2 = \beta_2 \frac{S_{y1}}{S_2} = 0{,}6026 \frac{5{,}8860}{0{,}1136} = 31{,}2227; \quad (90)$$

$$b_3 = \beta_3 \frac{S_{y1}}{S_3} = (-0{,}3944) \frac{5{,}8860}{0{,}0200} = -116{,}0719; \quad (91)$$

$$b_4 = \beta_4 \frac{S_{y1}}{S_4} = 0{,}3655 \frac{5{,}8860}{0{,}0500} = 43{,}0267. \quad (92)$$

Тоді

$$b_1 = \bar{y}_1 - \left[b_2\bar{x}_2 + b_3\bar{x}_3 + b_4\bar{x}_4\right] = 359{,}7079 - [31{,}2227 \cdot 6{,}34 - 116{,}0719 \cdot 0{,}30 + \\ + 43{,}0267 \cdot 0{,}22] = 187{,}1117. \quad (93)$$

Тоді модель (79) набуде вигляду:

$$\bar{y}_1(2,3,4) = 187{,}1117 + 31{,}2227x_2 - 116{,}0719x_3 + 43{,}0267x_4. \quad (94)$$

9. Рівняння залежности границі міцности алюмінієвого стопу АМг6 від концентрації основних компонентів у натуральній скалі та у натуральному позначенні має вигляд:

$$\sigma_в = 187{,}1117 + 31{,}2227 \cdot C(Mg) - 116{,}0719 \cdot C(Fe) + 43{,}0267 \cdot C(Si). \quad (95)$$

10. За формулою (74) множинний коефіцієнт кореляції:

$$r_1(2,3,4) = \sqrt{\beta_2 r_{12} + \beta_3 r_{13} + \beta_4 r_{14}} = \sqrt{0{,}6026 \cdot 0{,}5352 + (-0{,}3944) \cdot (-0{,}4273) + 0{,}3655 \cdot (-0{,}2659)} = \\ = 0{,}62758. \quad (96)$$

11. Коефіцієнт множинної кореляції після корекції – врахування числа параметрів рівняння (79) k = 4; числа експериментів N = 26 – за рівнянням (75) дорівнює:

$$\bar{r}_1(2,3,4) = \sqrt{1 - \left[1 - r_1^2(2,3,4)\right] \cdot \left(\frac{N-1}{N-k}\right)} = \sqrt{1 - \left[1 - 0,62758^2\right] \cdot \left(\frac{26-1}{26-4}\right)} = 0,55785. \quad (97)$$

12. Оцінка середньої квадратичної помилки під час розрахунку границі міцности зразка алюмінієвого стопу АМг6 за рівнянням моделі (70) у натуральній скалі за (78) становить:

$$\delta_{y1} = S(y_1)\sqrt{\left[1 - \bar{r}_1^2(2,3,4)\right] \cdot \left(\frac{N-1}{N-2}\right)} = 5,8860\sqrt{\left[1 - 0,55785^2\right] \cdot \left(\frac{26-1}{26-2}\right)} = 4,9858\ \text{МПа}. \quad (98)$$

13. Висунемо нульову гіпотезу H_0 відносно генерального множинного коефіцієнта кореляції $\rho_1(2,3,4)$, оцінкою якого є вибірковий множинний коефіцієнт кореляції $r_1(2,3,4)$, розрахований за формулою (74):

$$\left.\begin{array}{c} H_0\text{: } \rho_{1(2,3,4)} = 0 \\ \uparrow \\ r_{1(2,3,4)} \neq 0 \end{array}\right\} \quad (99)$$

14. Перевірка нульової гіпотези H_0 (99) За критичним значенням коефіцієнта кореляції, за критерієм Стьюдента та z-перетворенням Фішера привела до таких результатів:

1) за критичним значенням коефіцієнта кореляції [4, 5]:

для α=0,05

$$|r_1(2,3,4)| = 0,62758 > r_{кр.}\{q=1-0,05/2=0,975;\ f=N-2=24\} = 0,3882\ [5]; \quad (100)$$

- **для α=0,01**

$$|r_1(2,3,4)| = 0,62758 > r_{кр.}\{q=1-0,01/2=0,995;\ f=N-2=24\} = 0,4958\ [5], \quad (101)$$

тобто, з ймовірностями p=0,95 та p=0,99 при рівнях значущости α=0,05 та α=0,01 нульова гіпотеза відкидається. Тоді стверджуємо, що за $\mathbf{r_{кр.}}$ між границею міцности стопу АМг6 та концентраціями основних компонентів C(Mg), C(Fe) та C(Si) існує статистично надійний множинний лінійний зв'язок, при цьому **ступінь лінійности множинного зв'язку** дорівнює:

- **для α=0,05** $\xi_1(r)_{0,05} = \dfrac{|r_1(2,3,4)|}{r_{кр}(\alpha = 0,05)} = \dfrac{0,62758}{0,3882} = 1,6166 > 1;$

(102)

- **для α=0,01** $\xi_1(r)_{0,01} = \dfrac{|r_1(2,3,4)|}{r_{кр}(\alpha = 0,01)} = \dfrac{0,62758}{0,4958} = 1,2658 > 1,$

(103)

а **ступінь залишків нелінійности у лінійному множинному зв'язку** дорівнює:

- **для α=0,05** $\xi_2(r)_{0,05} = \dfrac{r_{кр}(\alpha = 0,05)}{|r_1(2,3,4)|} = \dfrac{0,3882}{0,62758} = 0,6186 < 1;$ (104)

• **для α=0,01** $\xi_2(r)_{0,01} = \dfrac{r_{кр}(\alpha = 0{,}01)}{|r_1(2,3,4)|} = \dfrac{0{,}4958}{0{,}62758} = 0{,}7900 < 1;$ (105)

загальна ступінь щільности множинного зв'язку (лінійного + нелінійного) за $\mathbf{r_{кр}}$.:

• **для α=0,05** $\xi_{12}(r)_{0,05} = \xi_1(r)_{0,05} + \xi_2(r)_{0,05} = 1{,}6166 + 0{,}6186 = 2{,}2352;$

(106)

• **для α=0,01** $\xi_{12}(r)_{0,01} = \xi_1(r)_{0,01} + \xi_2(r)_{0,01} = 1{,}2658 + 0{,}7900 = 2{,}0558;$

(107)

2) **за критерієм Стьюдента** [3, 5]:

• **для α=0,05**

$$|t_p| = \frac{|r_1(2,3,4)|}{\sqrt{1 - r_1^2(2,3,4)}}\sqrt{N-2} = \frac{0{,}62758}{\sqrt{1 - 0{,}62758^2}}\sqrt{26-2} = 3{,}9490 > t_T\,\{q = 1 - 0{,}05/2 = 0{,}975, \quad (108)$$
$$f = N - 2 = 26 - 2 = 24\} = 2{,}064\,[5];$$

• **для α=0,01** $|t_p| = 3{,}9490 > t_T$ {q=1-0,01/2=0,995; f=24}=2,797 [5], (109)

тобто, з ймовірностями p=0,95 та p=0,99 при рівнях значущости α=0,05 та α=0,01 нульова гіпотеза $\mathbf{H_0}$ відкидається. Тоді стверджуємо, що за t-критерієм між границею міцности стопу АМг6 та концентраціями основних компонентів C(Mg), C(Fe) та C(Si) існує статистично надійний множинний лінійний зв'язок, при цьому **ступінь лінійности множинного зв'язку** дорівнює:

• **для α=0,05** $\xi_1(t)_{0,05} = \dfrac{|t_p|}{t_т(\alpha = 0{,}05)} = \dfrac{3{,}9490}{2{,}064} = 1{,}9133 > 1;$ (110)

• **для α=0,01** $\xi_1(t)_{0,01} = \dfrac{|t_p|}{t_т(\alpha = 0{,}01)} = \dfrac{3{,}9490}{2{,}797} = 1{,}4119 > 1,$ (111)

а **ступінь залишків нелінійности у лінійному множинному зв'язку** дорівнює:

• **для α=0,05** $\xi_2(t)_{0,05} = \dfrac{t_т(\alpha = 0{,}05)}{|t_p|} = \dfrac{2{,}064}{3{,}9490} = 0{,}5227 < 1;$ (112)

• **для α=0,01** $\xi_2(t)_{0,01} = \dfrac{t_т(\alpha = 0{,}01)}{|t_p|} = \dfrac{2{,}797}{3{,}9490} = 0{,}7083 < 1;$ (113)

загальна ступінь щільности множинного зв'язку (лінійного+нелінійного) за t-критерієм Стьюдента:

для α=0,05 $\xi_{12}(t)_{0,05} = \xi_1(t)_{0,05} + \xi_2(t)_{0,05} = 1{,}9133 + 0{,}5227 = 2{,}4360;$ (114)

для α=0,01 $\xi_{12}(t)_{0,01} = \xi_1(t)_{0,01} + \xi_2(t)_{0,01}$=1,4119+0,7083=2,1202; (115)

3) за Z-перетворенням Фішера [3, 5]:

для α=0,05

$$\left|z_p\right| = \frac{1}{2}\ln\frac{1+r_1(2,3,4)}{1-r_1(2,3,4)} = \frac{1}{2}\ln\frac{1+0,62758}{1-0,62758} = 0,7374 > \left[(z_T\,\{q = 1 - \alpha/2 = 1 - 0,05/2 = \right.$$

$$\left. = 0,975\} = 1,96\,[24]) \cdot (\sigma_z = \frac{1}{\sqrt{N-3}} = \frac{1}{\sqrt{26-3}} = 0,2085) = 0,4087\right]; \quad (116)$$

для α=0,01 $\left|z_p\right| = 0,7374 > [(z_T\,\{q=0,995\}=2,58\,[24])\cdot(\sigma_Z=0,2085)=0,5379]$, (117)

тобто, з ймовірностями p=0,95 та p=0,99 при рівнях значущости α=0,05 та α=0,01 нульова гіпотеза $\mathbf{H_0}$ відкидається. Тоді стверджуємо, що за Z-перетворенням Фішера між границею міцности стопу АМг6 та концентраціями основних компонентів C(Mg), C(Fe) та C(Si) існує статистично надійний множинний лінійний зв'язок, при цьому **ступінь лінійности множинного зв'язку** дорівнює:

- **для α=0,05** $\xi_1(z)_{0,05} = \frac{\left|z_p\right|}{\left[z_т(q = 0,975) \cdot \sigma_z\right]} = \frac{0,7374}{0,4087} = 1,8043 > 1;$ (118)

- **для α=0,01** $\xi_1(z)_{0,01} = \frac{\left|z_p\right|}{\left[z_т(q = 0,995) \cdot \sigma_z\right]} = \frac{0,7374}{0,5379} = 1,3709 > 1,$ (119)

а **ступінь залишків нелінійности у лінійному множинному зв'язку** дорівнює:

- **для α=0,05** $\xi_2(z)_{0,05} = \frac{\left[z_т(q = 0,975) \cdot \sigma_z\right]}{\left|z_p\right|} = \frac{0,4087}{0,7374} = 0,5542 < 1;$ (120)

- **для α=0,01** $\xi_2(z)_{0,01} = \frac{\left[z_т(q = 0,995) \cdot \sigma_z\right]}{\left|z_p\right|} = \frac{0,5379}{0,7374} = 0,7295 < 1;$ (121)

загальна ступінь щільности множинного зв'язку (лінійного + нелінійного) за z-перетворенням Фішера:

- **для α=0,05** $\xi_{12}(z)_{0,05} = \xi_1(z)_{0,05} + \xi_2(z)_{0,05}$=1,8043+0,5542=2,3585; (122)

- **для α=0,01** $\xi_{12}(z)_{0,01} = \xi_1(z)_{0,01} + \xi_2(z)_{0,01}$=1,3709+0,7295=2,1004. (123)

15. Висунемо нульову гіпотезу $\mathbf{H_0}$ відносно ґенерального множинного коефіцієнта кореляції $\overline{\rho}_{1(2,3,4)}$, оцінкою якого є вибірковий множинний коефіцієнт кореляції $r_1(2,3,4)$, підданий **корекції** за рівнянням (75) з врахування числа параметрів рівняння (70): k=4 та числа спостережень N=26:

$$\left.\begin{array}{c} H_0': \ \bar{\rho}_{1(2,3,4)} = 0 \\ \uparrow \\ \bar{r}_{1(2,3,4)} \neq 0 \end{array}\right\} \qquad (124)$$

16. Здійснемо перевірку нульової гіпотези H_0' (124) за процедурою (100)–(123):

1) за критичним значенням коефіцієнта кореляції [4, 5]:

- **для α=0,05** $|\bar{r}_1(2,3,4)|=0{,}5579>r_{кр.}\{q=0{,}975;\ f=24\}=0{,}3882$ [5]; (125)
- **для α=0,01** $|\bar{r}_1(2,3,4)|=0{,}5579>r_{кр.}\{q=0{,}995;\ f=24\}=0{,}4958$ [5], (126)

тобто, з ймовірностями p=0,95 та p=0,99 H_0' відкидається. Тоді стверджуємо, що за $r_{кр.}$ між $\sigma_в$ та концентраціями основних компонентів C(Mg), C(Fe) та C(Si) існує статистично надійний множинний лінійний зв'язок після корекції, при цьому **ступінь лінійности зв'язку** дорівнює:

- **для α=0,05** $\xi_1(\bar{r})_{0,05}=\dfrac{|\bar{r}_1(2,3,4)|}{r_{кр}(\alpha=0{,}05)}=\dfrac{0{,}5579}{0{,}3882}=1{,}4372>1;$ (127)
- **для α=0,01** $\xi_1(\bar{r})_{0,01}=\dfrac{|\bar{r}_1(2,3,4)|}{r_{кр}(\alpha=0{,}01)}=\dfrac{0{,}5579}{0{,}4958}=1{,}1253>1,$ (128)

а **ступінь залишків нелінійности у лінійному множинному зв'язку** дорівнює:

- **для α=0,05** $\xi_2(\bar{r})_{0,05}=\dfrac{r_{кр}(\alpha=0{,}05)}{|\bar{r}_1(2,3,4)|}=\dfrac{0{,}3882}{0{,}5579}=0{,}6958<1;$ (129)
- **для α=0,01** $\xi_2(\bar{r})_{0,01}=\dfrac{r_{кр}(\alpha=0{,}01)}{|\bar{r}_1(2,3,4)|}=\dfrac{0{,}4958}{0{,}5579}=0{,}8887<1;$ (130)

загальна ступінь щільности множинного зв'язку (лінійного + нелінійного) за $r_{кр.}$:

- **для α=0,05** $\xi_{12}(r)_{0,05}=\xi_1(r)_{0,05}+\xi_2(r)_{0,05}=1{,}4372+0{,}6958=2{,}1330;$ (131)
- **для α=0,01** $\xi_{12}(r)_{0,01}=\xi_1(r)_{0,01}+\xi_2(r)_{0,01}=1{,}1253+0{,}8887=2{,}0140;$ (132)

2) за критерієм Стьюдента [3, 5]:

- **для α=0,05**

$$|t_p|=\frac{|r_1(2,3,4)|}{\sqrt{1-r_1^2(2,3,4)}}\sqrt{N-2}=\frac{0{,}55785}{\sqrt{1-0{,}55785^2}}\cdot\sqrt{26-2}=3{,}2929>t_T\{q=0{,}975,\ f=24\}=2{,}064\ [5]; \qquad (133)$$

- **для α=0,01** $|t_p|=3{,}2929>t_T\{q=0{,}995;\ f=24\}=2{,}797$ [5], (134) тобто, з ймовірностями p=0,95 та p=0,99 $\mathbf{H_0'}$ відкидається. Тоді стверджуємо, що за t-критерієм між $\sigma_в$ та концентраціями основних компонентів C(Mg), C(Fe) та C(Si) існує статистично надійний множинний лінійний зв'язок після корекції, при цьому **ступінь лінійности зв'язку дорівнює**:

- **для α=0,05** $\xi_1(t)_{0,05} = \dfrac{|t_p|}{t_т(\alpha = 0,05)} = \dfrac{3,2929}{2,064} = 1,5954 > 1;$ (135)

- **для α=0,01** $\xi_1(t)_{0,01} = \dfrac{|t_p|}{t_т(\alpha = 0,01)} = \dfrac{3,2929}{2,797} = 1,1773 > 1,$ (136)

а **ступінь залишків нелінійности у лінійному множинному зв'язку** дорівнює:

- **для α=0,05** $\xi_2(t)_{0,05} = \dfrac{t_т(\alpha = 0,05)}{|t_p|} = \dfrac{2,064}{3,2929} = 0,6268 < 1;$ (137)

- **для α=0,01** $\xi_2(t)_{0,01} = \dfrac{t_т(\alpha = 0,01)}{|t_p|} = \dfrac{2,797}{3,2929} = 0,8494 < 1;$ (138)

загальна ступінь щільности множинного зв'язку (лінійного + нелінійного) за t-критерієм Стьюдента:

- **для α=0,05** $\xi_{12}(t)_{0,05} = \xi_1(t)_{0,05} + \xi_2(t)_{0,05} = 1,5954 + 0,6268 = 2,2222;$ (139)

- **для α=0,01** $\xi_{12}(t)_{0,01} = \xi_1(t)_{0,01} + \xi_2(t)_{0,01} = 1,1773 + 0,8494 = 2,0267;$ (140)

3) за Z-перетворенням Фішера [3, 5]:

- **для α=0,05**

$$|z_p| = \frac{1}{2}\ln\frac{1+\bar{r}_1(2,3,4)}{1-\bar{r}_1(2,3,4)} = \frac{1}{2}\ln\frac{1+0,55785}{1-0,55785} = 0,6297 >$$
$$> \left[(z_T\{q = 0,975\} = 1,96\,[24])(\sigma_z = 0,2085) = 0,4087\right]; \quad (141)$$

- **для α=0,01** $|z_p| = 0,6297 > [(z_T\{q=0,995\}=2,58\,[24])\cdot(\sigma_Z=0,2085) =$

$0,5379],$ (142)

тобто, з ймовірностями p=0,95 та p=0,99 $\mathbf{H_0'}$ відкидається. Тоді стверджуємо, що за Z-перетворенням Фішера між $\sigma_в$ та концентраціями основних компонентів C(Mg), C(Fe) та C(Si) існує статистично надійний множинний лінійний зв'язок після корекції, при цьому **ступінь лінійности зв'язку** дорівнює:

- **для α=0,05** $\xi_1(z)_{0,05} = \dfrac{|z_p|}{[z_т(q = 0,975)\cdot\sigma_z]} = \dfrac{0,6297}{0,4087} = 1,5407 > 1;$ (143)

- **для α=0,01** $\xi_1(z)_{0,01} = \dfrac{|z_p|}{[z_т(q = 0,995)\cdot\sigma_z]} = \dfrac{0,6297}{0,5379} = 1,1707 > 1,$ (144)

а **ступінь залишків нелінійности у лінійному множинному зв'язку** дорівнює:

- **для α=0,05** $\xi_2(z)_{0,05} = \frac{[z_т(q=0{,}975)\cdot\sigma_z]}{|z_p|} = \frac{0{,}4087}{0{,}6297} = 0{,}6491 < 1;$ (145)

- **для α=0,01** $\xi_2(z)_{0,01} = \frac{[z_т(q=0{,}995)\cdot\sigma_z]}{|z_p|} = \frac{0{,}5379}{0{,}6297} = 0{,}8542 < 1;$ (146)

загальна ступінь щільности множинного зв'язку (лінійного + нелінійного) за z-перетворенням Фішера:

- **для α=0,05** $\xi_{12}(z)_{0,05} = \xi_1(z)_{0,05} + \xi_2(z)_{0,05} = 1{,}5407 + 0{,}6491 = 2{,}1898;$ (147)
- **для α=0,01** $\xi_{12}(z)_{0,01} = \xi_1(z)_{0,01} + \xi_2(z)_{0,01} = 1{,}1707 + 0{,}8542 = 2{,}0249.$ (148)

Підсумкова матриця значень множинних коефіцієнтів кореляцій, результати перевірки їх значущости за критеріями $\mathbf{r_{кр}}$, t-Стьюдента та F-перетворення Фішера (Z) та ступенів лінійности і нелінійности приведена у табл. 7.

Таблиця 7

Підсумкова матриця значень множинних коефіцієнтів лінійних кореляцій та ступенів лінійности і нелінійности за критеріями $r_{кр}$, t-Стьюдента та F-перетворення Фішера (Z)

За критерієм	До корекції		Після корекції	
	$r_1(2,3,4) = 0{,}62758$		$\bar{r}_1(2,3,4) = 0{,}55785$	
критичним коефіцієнтом	α=0,05	α=0,01	α=0,05	α=0,01
ступінь лінійности $\xi_1(r)$	1,6166	1,2658	1,4372	1,1253
ступінь нелінійности $\xi_2(r)$	0,6186	0,7900	0,6958	0,8887
загальна ступінь щільности зв'язку	2,2352	2,0558	2,1330	2,0140
Стьюдента (t)	α=0,05	α=0,01	α=0,05	α=0,01
ступінь лінійности $\xi_1(t)$	1,9133	1,4119	1,5954	1,1773
ступінь нелінійности $\xi_2(t)$	0,5227	0,7083	0,6268	0,8494
загальна ступінь щільности зв'язку	2,4360	2,1202	2,2222	2,0267
F-перетворенням Фішера (Z)	α=0,05	α=0,01	α=0,05	α=0,01
Ступінь лінійности $\xi_1(z)$	1,8043	1,3709	1,5407	1,1707
Ступінь нелінійности $\xi_2(z)$	0,5542	0,7295	0,6491	0,8542
загальна ступінь щільности зв'язку	2,3585	2,1004	2,1898	2,0249

3. Нелінійний кореляційний зв'язок. Кореляційні співвідношення

У разі прийняття нульової гіпотези $H_0 : \rho=0$ (рівности нулю коефіцієнта кореляції) та оцінки ступеня нелінійности за $\xi_2 > 1$ за критеріями $r_{кр}$, t, $(Z_T \cdot \sigma_Z)$ доводиться існування нелінійного зв'язка між **X** та **Y**, щільности якому можна дати оцінку за кореляційними співвідношеннями $\mathbf{R^2}$ і **R**:

$$R^2 = 1 - \frac{SS_{ад}}{SS_{від}}; \tag{149}$$

$$R = +\sqrt{1 - \frac{S^2_{ад}}{S^2_{від}}}, \tag{150}$$

де $$S^2_{ад} = \frac{SS_{ад}}{f_{ад}} = \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N - \lambda} \tag{151}$$

дисперсія адекватности апроксимаційної моделі;
N – число точок плану експеримента;

$SS_{ад}$ – сума квадратів (чисельник) відхилення i-результату експерименту від i-значення, розрахованого за апроксимаційною моделлю (чисельник);

$f_{ад}$ – число ступенів вільностей під час розрахунку дисперсій адекватности (знаменник);
λ – число коефіцієнтів математичної моделі;

$$S^2_{від} = \frac{SS_{від}}{f_{від}} = \frac{\sum_{i=1}^{N}(y_{ij} - \bar{y}_i)^2}{N - 1} \tag{152}$$

дисперсія відновлення (помилквсього експерименту);

$SS_{від}$ – сума квадратів відхилення ij-результату від i-середньої (чисельник);

$f_{від}$ – число ступенів вільностей під час розрахунку дисперсій відновлення (знаменник);R^2 та R можуть набувати таких значень:

$$0 \le R^2 \le 1; \ 0 \le R \le 1. \tag{153}$$

Оцінку ступеню нелінійности зв'язку можна дати за співвідношеннями:

$$\xi_1(R^2) = \frac{R^2}{r}; \ \ \xi_1(R) = \frac{R}{r}, \tag{154}$$

а оцінку ступеню лінійности зв'язку – за співвідношеннями:

$$\xi_2(R^2) = \frac{r}{R^2}; \ \ \xi_2(R) = \frac{r}{R}. \tag{155}$$

Співвідношення:

$$\xi_0(R) = \frac{R^2}{R} \tag{156}$$

може вказати на ступінь чистої адекватности (чистої апроксимації).

Список використаних літературних джерел інформації

1. Лариков Л.Н. Структура и свойства металлов и сплавов. Справочник: Тепловые свойства металлов и сплавов / Лариков Л.Н., Юрченко Ю.Ф. − Киев: Наукова думка, 1985. − 438 с.
2. Сіренко Г.О., Базюк Л.В., Мещерякова Н.В. Теплофізичні властивости металів та стопів: залежність зміни ентальпії від температури та радіусу атомів // Фізика і хімія твердого тіла. – Т. 12, № 1. – 2011. – С. 197-207.
3. Степнов М.Н. Статистическая обработка результатов механических испытаний / Михаил Никитович Степнов. – Москва: Машиностроение, 1972. – 232 с. : ил., табл. – Библиогр.: с. 229-230 (36 назв.).
4. Адлер Ю.П. Планирование эксперимента при поиске оптимальных условий / Ю.П. Адлер, Е.В. Маркова, Ю.В. Грановский. – 2-е изд., перераб. и допол. – Москва: Наука, 1976. – 280 с.: ил., табл. – Библиогр. в конце гл.
5. Мюллер П., Нойман П., Шторм Р. Таблицы по математической статистике / Пер. с нем. и предисловие В.М. Ивановой.– Москва: Финансы и статистика, 1982. – 272 с.: ил.
6. Сіренко Г.О., Мідак Л.Я., Сіренко О.Г. Методи лінійної множинної кореляції та регресії в хемічному матеріалознавстві // Вісник Прикарп. нац. ун-ту ім. Василя Стефаника. Серія Хімія. – Івано-Франківськ: 2011. – Вип. XII. – С. 124-132.
7. Володарський Є.Т., Кошева Л.О. Статистична обробка даних: Навчальний посібник. – Київ: Нац. авіац. ун-т, 2008. – 308 с. – ISBN 978-966-598-406-1.
8. Неділько С.А. Математичні методи в хімії: Підручник. – Київ: Либідь, 2005. – 256 с. – ISBN 966-06-0384-3.
9. Михайліченко Б.М. Курс загальної хімії. Теоретичні основи: Навчальний посібник. – Київ: Вид. «Знання» 2009. – 549с. ISBN 978-966-346-712-2.
10. Математическая энциклопедия у 2-х том./Ред. колегія: С.И. Адян, П.С. Александров, Н.С. Бахвалов и др. / Глав. ред. И.М. Виноградов. – Москва: Изд. «Сов. энциклопедия», Т.1-1977. – 1151с.; Т.2. – 1979. – 1103с.

Сіренко Г.О. – професор, доктор технічних наук, професор кафедри хімії;
Складанюк М.Б. –к.ф.-м.н., доцент кафедри хімії;
Мартинюк М. І. – магістр, здобувач наукової ступені по кафедрі хімії;
Говдяк І.В. – студентка IV курсу бакалаврату.