Лекція №7.

План.

Кореляційний і регресійний методи аналізу зв'язку

Основне завдання кореляційного і регресійного аналізу статистичних даних є виявлення залежності між досліджуваними ознаками у вигляді певної математичної формули і встановлення за допомогою коефіцієнта кореляції порівняльної ознаки тісноти взаємозв'язку.

Кореляційний і регресійний методи аналізу розв'язують два основних завдання:

1)визначають з допомогою рівняння регресії аналітичну форму зв'язку між варіацією ознак «х» і «у»;
2)встановлюють міру тісноти зв'язку між ознаками.

В практиці економіко-статистичних досліджень часто доводиться мати справу з прямолінійною формою зв'язку, яка виражається за допомогою рівняння регресії.

Рівняння регресії характеризує зміну середнього рівняння результативної ознаки «у» в залежності від зміни факторної ознаки «х».
У випадку лінійної форми зв'язку рівняння регресії має вигляд:

$$\widetilde{y_x} = a_0 + a_1 x ,$$

де $\widetilde{y_x}$ – вирівняне середнє значення результативної ознаки;

х – значення факторної ознаки;

a_0 і a_1 – параметри рівняння;

a_0 – значення «у» при х=0;

a_1 – коефіцієнт регресії.

Коефіцієнт регресії « a_1» показує наскільки зміниться результативна ознака «у» при зміні факторної ознаки «х» на одиницю.

Якщо « a_1» має позитивний знак, то зв'язок прямий, якщо від'ємний – зв'язок обернений.

Параметри рівняння зв'язку визначаються способом найменших квадратів .

Повернутись до плану.

Метод найменших квадратів

У процесі вивчення різних питань природознавства, економіки і техніки, соціології, педагогіки доводиться на основі великої кількості дослідних даних виявляти суттєві фактори, які впливають на досліджуваний об'єкт, а також встановлювати форму зв'язку між різними зв'язаними одна з одною величинами (ознаками).

Нехай у результаті досліджень дістали таку таблицю деякої функціональної залежності:

Таблиця 1

x	x1	x2	…	xn
y	y1	y2	…	yn

Треба знайти аналітичний вигляд функції $y = f(x)$, яка добре відображала б цю таблицю дослідних даних. Функцію $y = f(x)$ можна шукати у вигляді інтерполяційного поліному. Але інтерполяційні поліноми не завжди добре відображають характер поведінки таблично заданої функції. До того ж значення y_1 дістають у результаті експерименту, а вони, як правило, сумнівні. У цьому разі задача інтерполювання табличної функції втрачає

сенс. Тому шукають таку функцію $y = F(x)$, значення якої при $x = x_i$ досить близькі до табличних значень y_i $(i = 1, 2, \ldots, n)$. Формулу $y = F(x)$ називають емпіричною, або рівнянням регресії y від x. Емпіричні формули мають велике практичне значення, вдало підібрана емпірична формула дає змогу не тільки апроксимувати сукупність експериментальних даних, «згладжуючи» значення величини y, а й екстраполювати знайдену залежність на інші проміжки значень x.

Процес побудови емпіричних формул складається з двох етапів: встановлення загального виду цієї формули і визначення найкращих її параметрів.

Щоб встановити вигляд емпіричної формули, на площині будують точки з координатами (x_i, y_i) $(i = 1, 2, \ldots, n)$. Деякі з цих точок сполучають плавною кривою, яку проводять так, щоб вона проходила якомога ближче до всіх даних точок. Після цього візуально визначають, графік якої з відомих нам функцій найкраще підходить до побудованої кривої. Звичайно, намагаються підібрати найпростіші функції: лінійну, квадратичну, дробово-раціональну, степеневу, показникову, логарифмічну.

Встановивши вигляд емпіричної формули, треба знайти її параметри (коефіцієнти). Найточніші значення коефіцієнтів емпіричної формули визначають методом найменших квадратів. Цей метод запропонували відомі математики К. Гаусс і А. Лежандр.

Розглянемо суть методу найменших квадратів.

Нехай емпірична формула має вигляд

$$y = F(x; a_1, a_2, \ldots, a_m), \qquad (1)$$

де $a_1, a_2, \ldots a_m$ — невідомі коефіцієнти. Треба знайти такі значення коефіцієнтів a_i $(i = 1,2,\ldots, m)$, за яких крива (1) якомога ближче проходитиме до

всіх n точок $(x_1, y_1),(x_2, y_2),...,(x_n, y_n)$, знайдених експериментально. Зрозуміло, що жодна з експериментальних точок не задовольняє точно рівняння (1). Відхилення від підстановки координат (x_i, y_i) у рівняння (1) дорівнюватимуть величинам

$$\delta_i = y_i - F(x_i; a_1, \ldots, a_m) \quad (i = 1, 2, \ldots, n).$$

За методом найменших квадратів найкращі значення коефіцієнтів $a_1, a_2, \ldots a_m$ ті, для яких сума квадратів відхилень дослідних даних y_i від обчислених за емпіричною формулою (1) найменша

$$S(a_1, a_2, \ldots, a_m) = \sum_{i=1}^{n} \delta_i^2 = \sum_{i=1}^{n} \left(y_i - F(x_i; a_1, \ldots, a_m)\right)^2 \qquad (2)$$

Звідси випливає, що величина (2), яка є функцією від коефіцієнтів $a_1, a_2, \ldots a_m$, повинна мати мінімум. Необхідна умова мінімуму функції багатьох змінних — її частинні похідні мають дорівнювати нулю, тобто

$$\frac{\partial S}{\partial a_1} = 0, \quad \frac{\partial S}{\partial a_2} = 0, \ldots, \quad \frac{\partial S}{\partial a_m} = 0.$$

Диференціюючи вираз (2) по невідомих параметрах $a_1, a_2, \ldots a_m$, матимемо відносно них систему рівнянь:

$$\begin{cases} \sum_{i=1}^{n} \left(y_i - F(x_i; a_1, a_2, \ldots, a_m)\right) \cdot \dfrac{\partial F(x_i; a_1, \ldots, a_m)}{\partial a_1} = 0, \\ \sum_{i=1}^{n} \left(y_i - F(x_i; a_1, \ldots, a_m)\right) \cdot \dfrac{\partial F(x_i; a_1, \ldots, a_m)}{\partial a_2} = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \sum_{i=1}^{n} \left(y_i - F(x_i; a_1, \ldots, a_m)\right) \cdot \dfrac{\partial F(x_i; a_1, \ldots, a_m)}{\partial a_m} = 0. \end{cases}$$

Система (3) називається нормальною. Якщо вона має розв'язок, то він єдиний, і буде шуканим.

Якщо емпірична функція (1) лінійна відносно параметрів $à_1, à_2, ... a_m$, то нормальна система (3) буде системою з *m* лінійних рівнянь відносно шуканих параметрів.

Будуючи емпіричні формули, припускатимемо, що експериментальні дані $(x_i, y_i)\ i = 1,2,..., n$ додатні.

Якщо серед значень x_i і y_i є від'ємні, то завжди можна знайти такі додатні числа p і q, що $\bar{x}_i = x_i + p > 0$ і $\bar{y}_i = y_i + q > 0 \quad (i = 1, 2, \ldots, n)$.

Тому розв'язування поставленої задачі завжди можна звести до побудови емпіричної формули для додатних значень $(\bar{x}_i, \bar{y}_i)$.

3.Побудова лінійної емпіричної формули.

Нехай між даними $(x_i, y_i) \quad (i = 1, 2, \ldots, n)$ існує лінійна залежність. Шукатимемо емпіричну формулу у вигляді

$$y = ax + b, \qquad (4)$$

де коефіцієнти a і b невідомі.

Знайдемо значення a і b, за яких функція $S(a, b) = \sum_{i=1}^{n} (y_i - ax_i - b)^2$ матиме мінімальне значення. Щоб знайти ці значення, прирівняємо до нуля частинні похідні функції

$$\begin{cases} \dfrac{\partial S}{\partial a} = 2\sum_{i=1}^{n} (y_i - ax_i - b)(-x_i) = 0, \\ \dfrac{\partial S}{\partial b} = 2\sum_{i=1}^{n} (y_i - ax_i - b)(-1) = 0. \end{cases}$$

Звідси, врахувавши, що

$$\sum_{i=1}^{n} b = nb,$$

маємо

$$\begin{cases} a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i, \\ a\sum_{i=1}^{n} x_i + nb = \sum_{i=1}^{n} y_i. \end{cases} \qquad (5)$$

Розв'язавши відносно a і b останню систему, знайдемо

$$a = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}, \qquad (6)$$

$$b = \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i y_i \sum_{i=1}^{n} x_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}. \qquad (7)$$

Зазначимо, що, крім графічного, є ще й аналітичний критерій виявлення лінійної залежності між значеннями x і y.

Покладемо $\Delta x_i = x_{i+1} - x_i$, $\Delta y_i = y_{i+1} - y_i$,

$k_i = \Delta y_i / \Delta x_i \quad (i = 1, 2, \ldots, n-1)$.

Якщо $k_i = const$, то залежність між x і y лінійна, бо точки (x_i, y_i) лежатимуть на одній прямій. Якщо $k_1 \approx k_2 \approx \ldots \approx k_{n-1}$, то між x і y існує майже лінійна залежність, оскільки точки (x_i, y_i) лежатимуть близько до деякої прямої.

Розрахунок параметрів лінійного рівняння зв'язку і лінійного коефіцієнта кореляції між вартістю основних виробничих фондів і випуском продукції.

Номер заводу (n)	Вартість основних виробничих фондів, млн.грн. (х)	Випуск продукції, млн.грн. (у)	x^2	ху	y^2	$\widetilde{y_x} = a_0 + a$
1	12	5,6	144	67,2	31,36	5,2
2	8	4,0	64	32,0	16,00	3,5
3	10	4,0	100	40,0	16,00	4,4
4	6	2,4	36	14,4	5,76	2,7
5	9	3,6	81	32,4	12,96	4,0
6	15	5,0	225	75,0	25,00	6,5
7	11	4,6	121	50,6	21,16	4,8
8	13	6,5	169	84,5	42,25	5,6
9	14	7,0	196	98,0	49,00	6,1
10	10	4,5	100	45,0	20,25	4,4
Разом:	108	47,2	1236	539,1	239,74	47,2
В середньому на 1 завод	10,8	4,72	132,6	53,91	23,974	4,72

За способом найменших квадратів визначаємо параметри:

$$a_0 = \frac{1236 \cdot 47{,}2 - 108 \cdot 596{,}1}{10 \cdot 1236 - 108 \cdot 108} = \frac{58339{,}2 - 58222{,}8}{12360 - 11664} = \frac{116{,}4}{696{,}0} = 0{,}167;$$

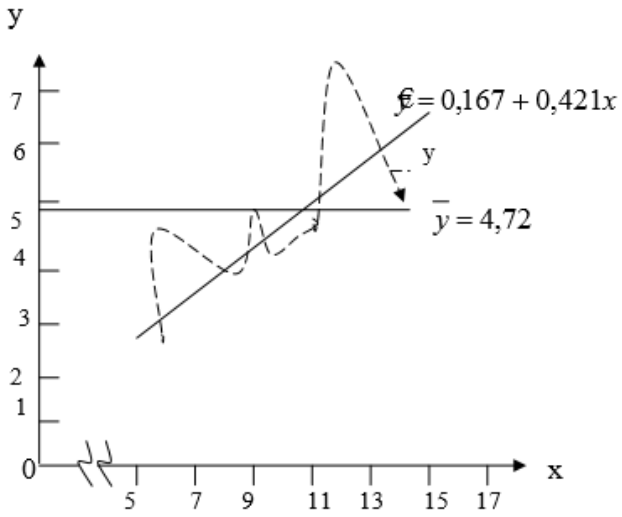
$$a_1 = \frac{10 \cdot 539{,}1 - 108 \cdot 47{,}2}{696{,}0} = \frac{5391{,}0 - 5097{,}6}{696{,}0} = \frac{293{,}4}{696{,}0} = 0{,}421.$$

Лінійне рівняння регресії між вартістю основних виробничих фондів і випуском продукції має вигляд: $\widetilde{y_x} = 0{,}167 + 0{,}421x$.

Таким чином, при збільшенні вартості основних виробничих фондів на 1 млн.грн. випуск продукції зросте на 0,421 млн.грн.

Підставляючи в дане рівняння послідовно значення факторної ознаки «х», отримаємо вирівняні значення результативної ознаки «у». Якщо параметри рівняння визначені правильно, то $\sum y = \sum \widetilde{y_x} = 47{,}2$.

Побудуємо графік, який покаже вирівнювання емпіричних даних рівнянням прямої.



Повернутись до плану.

Для економічної інтерпретації лінійних і нелінійних зв'язків між двома досліджуваними явищами часто використовують розраховані на основі рівнянь регресії коефіцієнти еластичності.

Коефіцієнт еластичності показує, на скільки відсотків змінюється в середньому результативна ознака «у» при зміні факторної ознаки «х» на 1%. Для лінійної залежності коефіцієнт еластичності визначається за формулою:

$$\varepsilon = a_1 \frac{x}{\widetilde{y_x}}, \quad \varepsilon = a_1 \frac{\overline{x}}{\overline{y}}$$

де $\boldsymbol{\varepsilon}$ – коефіцієнт еластичності.

В нашому прикладі коефіцієнт еластичності на першому підприємстві при (х=12) буде дорівнювати: $\varepsilon_1 = a_1 \frac{x}{\widetilde{y_x}} = 0.421 \frac{12}{5.2} = 0.97\%$.

Отже, на 1 % приросту вартості основних виробничих фондів, випуск продукції зросте на 0,97 %. На п'ятому підприємстві при (х=9) $\varepsilon_5 = 0.421 \frac{9}{4} = 0.95\%$, на десятому – при (х=10) $\varepsilon_{10} = 0.421 \frac{10}{4{,}4} = 0.96\%$

Для всіх підприємств разом коефіцієнт еластичності становить:

$$\varepsilon = a_1 \frac{\overline{x}}{\overline{y}} = 0{,}421 \frac{10{,}8}{4.72} = 0{,}963\%$$

Це означає, що при збільшенні середньої вартості основних виробничих фондів на 1 % випуск продукції зросте в середньому на 0,963 %.

Якщо залежність між ознаки параболічна, то коефіцієнт еластичності визначається за формулою

$$\varepsilon = (a_1 + a_2 x) \frac{\overline{x}}{\overline{y}}$$

Визначення тісноти зв'язку в кореляційно-регресійному аналізі ґрунтується на правилі складання дисперсій, але для оцінки лінії регресії використовують теоретичні значення результативної ознаки.

Різниця між загальною і залишковою дисперсіями дає нам теоретичну (факторну) дисперсію, яка вимірює варіацію, зумовлену фактором «х». На порівнянні цієї різниці із загальною дисперсією побудований **індекс кореляції**,

або **теоретичне кореляційне відношення**, які обчислюються за формулами:

$$R = \sqrt{\frac{\sigma_з^2 - \sigma_e^2}{\sigma_з^2}} = \sqrt{1 - \frac{\sigma_e^2}{\sigma_з^2}}, \qquad або \qquad R = \sqrt{\frac{\delta^2}{\sigma_з^2}},$$

де R – індекс кореляції (теоретичне кореляційне відношення);

$\sigma_з^2$ – загальна дисперсія;

σ_e^2 – залишкова дисперсія;

σ^2 – факторна (теоретична) дисперсія.

Факторну дисперсію з теоретичних значень обчислюють за формулою:

$$\sigma^2 = \frac{\Sigma\left(\widetilde{y_x} - \overline{y}\right)^2}{n}$$

або за формулою без теоретичних значень:

$$\sigma^2 = \frac{(a_0 \sum y + a_1 \sum xy) - (\overline{y})^2}{n}$$

Залишкову дисперсію визначають за формулою:

$\sigma_e^2 = \frac{\sum(y-\overline{y_x})^2}{n}$

або за правилом складання дисперсій:

$$\sigma_e^2 = \sigma_з^2 - \sigma^2$$

В нашому прикладі факторна дисперсія дорівнює:

$$\delta^2 = \frac{(0{,}167 \cdot 47{,}2 + 0{,}421 \cdot 539{,}1) - 4{,}72^2}{10} = 1{,}206.$$

Загальна дисперсія становить:

$$\sigma_з^2 = \overline{y^2} - \left(\overline{y}\right)^2 = 23{,}974 - 22{,}278 = 1{,}696.$$

Залишкову дисперсію визначаємо як різницю між загальною і факторною дисперсіями:

$$\sigma_e^2 = \sigma_з^2 - \sigma^2 = 1{,}699 - 1{,}206 = 0{,}490$$

Таким чином індекс кореляції за вищенаведеними формулами буде дорівнювати:

$$R = \sqrt{\frac{\sigma_з^2 - \sigma_e^2}{\sigma_з^2}} = \sqrt{\frac{1{,}696 - 0{,}490}{1{,}696}} = 0{,}843,$$

або $$R = \sqrt{1 - \frac{\sigma_e^2}{\sigma_з^2}} = \sqrt{1 - \frac{0{,}490}{1{,}96}} = 0{,}843,$$

або $$R = \sqrt{\frac{\delta^2}{\sigma_з^2}} = \sqrt{\frac{1{,}206}{1{,}696}} = \sqrt{0{,}711} = 0{,}843.$$

Індекс кореляції показує тісну залежність випуску продукції від вартості основних фондів.

Повернутись до плану.

Коефіцієнт детермінації (R^2) характеризує ту частину варіації результативної ознаки «у», яка відповідає лінійному рівнянню регресії:

$$R^2 = \frac{\delta^2}{\sigma_з^2} = \frac{1{,}206}{1{,}696} = 0{,}711.$$

Отже, в обстеженій сукупності заводів 71,1 % варіації випуску продукції пояснюється різними рівнями оснащеності заводів основними фондами.

Індекс кореляції приймає значення від «0» до «1». Коли R=0, то зв'язку між варіацією ознак «у» і «х» немає. Залишкова дисперсія дорівнює загальній
($\sigma_e^2 = \sigma_з^2$), а теоретична дисперсія дорівнює нулю ($\sigma^2 = 0$).

При R=1 теоретична дисперсія дорівнює загальній ($\sigma^2 = \sigma_з^2$), а залишкова $\sigma_e^2 = 0$

Для вимірювання тісноти зв'язку і визначення його напрямку при лінійній залежності використовують **лінійний коефіцієнт кореляції**, який визначається за формулою:

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y}.$$

Значення «r» коливається в межах від -1до +1. Додатнє значення «r» означає прямий зв'язок між ознаками, а від'ємне – зворотній.
Оцінка тісноти зв'язку проводиться за схемою:

Сила зв'язку	Величина лінійного коефіцієнта кореляції при наявності:	
	прямого зв'язку	оберненого зв'язку
Слабка	0,1 – 0,30	(-0,1) – (-0,30)
Середня	0,3 – 0,70	(-0,3) – (-0,70)
Тісна	0,7 – 0,99	(-0,7) – (-0,99)

За даними нашого прикладу обчислимо лінійний коефіцієнт кореляції:
За даними нашого прикладу обчислимо лінійний коефіцієнт кореляції:

$$\sigma_x = \sqrt{\overline{x^2} - \left(\bar{x}\right)^2} = \sqrt{123,6 - 10,8^2} = \sqrt{6,96} = 2,638;$$

$$\sigma_y = \sqrt{\overline{y^2} - \left(\bar{y}\right)^2} = \sqrt{23,974 - 4,72^2} = 1,302;$$

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y} = \frac{53,91 - 10,8 \cdot 4,72}{2,638 \cdot 1,302} = \frac{2,9340}{3,4349} = 0,854.$$

Це означає, що зв'язок між вартістю основних виробничих фондів і випуском продукції сильний (тісний) і прямий.

Істотність зв'язку коефіцієнта детермінації R_2 перевіряють за допомогою таблиці критерію F для 5%-горівня значимості. Так, при

k_2=n-m=10-2=8.

Фактичне значення F- критерія для нашого прикладу визначають за формулою:

$$F_{\phi} = \frac{R^2}{1 - R^2} \cdot \frac{k_2}{k_1} = \frac{0,711}{1 - 0,711} \cdot \frac{8}{1} = 19,68.$$

Критичне значення $F_{т\,(0,95)}$ значно менше від фактичного

$F_{т\,(0,95)} < F_{ф}$ (5,32 < 19,68), то підтверджує істотність кореляційного зв’язку між досліджуваними ознаками.

Для встановлення достовірності обчисленого лінійного коефіцієнта кореляції використовують критерій Стьюдента (t- критерій):

$$t_r = \frac{|r|}{\mu_r},$$

де μ_r – середня помилка коефіцієнта кореляції, яку визначають за формулою:

$$\mu_r = \frac{1-r^2}{\sqrt{n-1}}.$$

При достатньо великому числі спостережень ($n > 50$) коефіцієнт кореляції можна вважати достовірним, якщо він перевищує свою помилку в 3 і більше раз, а якщо він менше 3, то зв'язок між досліджуваними ознаками «у» і «х» не доведений.

В нашому прикладі середня помилка коефіцієнта кореляції дорівнює:

$$\mu_r = \frac{1-0{,}853^2}{\sqrt{9}} = \frac{1-0{,}723}{3} = \frac{0{,}277}{3} = 0{,}092.$$

Відношення коефіцієнта кореляції до його середньої помилки становить:

t_r = 0,853/0,092 = 9,27.

Це дає нам право вважати, що обчислений лінійний коефіцієнт кореляції достатньо точно характеризує силу зв’язку між досліджуваними ознаками.

Повернутись до плану.

Критерій лінійності та коефіцієнт кореляції .

Після визначення вибору математичного апарату безпосередньо переходять до побудови математичної моделі. Якщо система підлягає опису алгебраїчними методами, то ця функціональна залежність матиме вигляд алгебраїчного рівняння. Але перш ніж приступити до підбору типу рівняння та визначення його параметрів необхідно визначити лінійність кореляції між досліджуваними величинами.

Створення математичної моделі, яка б описувала залежність між досліджуваними величинами починається з аналізу експериментальних даних, які отримані в результаті проведення досліду, а саме, з визначення типу кореляції між ними. Вірне визначення типу кореляційного зв'язку між ознаками дозволить підібрати відповідний вид функціональної залежності, якій підпорядковується взаємозв'язок між ними.

Для визначення ступеня наближення залежності між ознаками до лінійної чи нелінійної використовується критерій лінійності, який визначається за формулою

$$L = \eta^2 - r^2, \qquad (8)$$

де η - кореляційне відношення;

r - коефіцієнт кореляції.

Якщо критерій лінійності дорівнює нулю (L = 0), тобто $\eta^2 = r^2$, то взаємозв'язок між ознаками набуває лінійного характеру. Чим ближче значення критерію лінійності наближується до 1, тим чіткіше вимальовується нелінійний характер зв'язку між ознаками.

Коефіцієнт кореляції та кореляційне відношення використовуються для визначення тісноти зв'язку між ознаками. Але коефіцієнт кореляції застосовується для визначення тісноти зв'язку при лінійній залежності між ознаками. Цей коефіцієнт може змінюватись у межах: $-1 < r < +1$. Чим більше коефіцієнт кореляції наближується до 1, тим тісніший лінійний кореляційний зв'язок між ознаками. У випадку коли $r = 0$, між ознаками немає лінійного зв'язку, але нелінійна залежність може існувати. При нелінійній залежності між ознаками, для визначення тісноти зв'язку між ними, застосовується кореляційне відношення. Кореляційне відношення може приймати значення в межах $0 < \eta < 1$. Чим сильніший кореляційний зв'язок, тим більше значення η. При відсутності кореляції - $\eta = 0$.

Коефіцієнт кореляції визначається за формулою

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x}) \cdot (y_{\tilde{n}i} - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \cdot \sum_{i=1}^{n}(y_{\tilde{n}i} - \overline{y})^2}}, \qquad (9)$$

де x_i - вхідні величини;

y_i - вихідні величини;

$\overline{x} = \frac{\sum_{i=1} x_i}{n}$ - середнє арифметичне;

$\overline{y} = \frac{\sum_{i=1} y_i}{n}$ - середнє арифметичне;

n – кількість експериментальних точок.

Кореляційне відношення визначається за формулою

$$\eta = \sqrt{\frac{\sum_{i=1}^{n} n_i (y_{c_i} - \overline{y})^2}{\sum_{j=1}^{n} (\acute{o}_j - \overline{\acute{o}})^2}}, \qquad (10)$$

де y_{c_i} - умовні (групові) середні значення

Визначення критерію лінійності, візуальна оцінки характеру розташування точок на полі кореляції, досвід попередніх досліджень та міркувань, що базуються на знанні фізичної сутності процесу дає змогу визначити тип залежності між досліджуваними ознаками.

Залежність між двома ознаками може бути виражена у вигляді прямої або кривої лінії функціональної залежності, які в загальному вигляді представлені на рис. 1.

Розміщуючи на графіку експериментальні точки можна орієнтовно визначити тип лінії, а відповідно і тип функціональної залежності, яка відповідає залежності між ознаками.

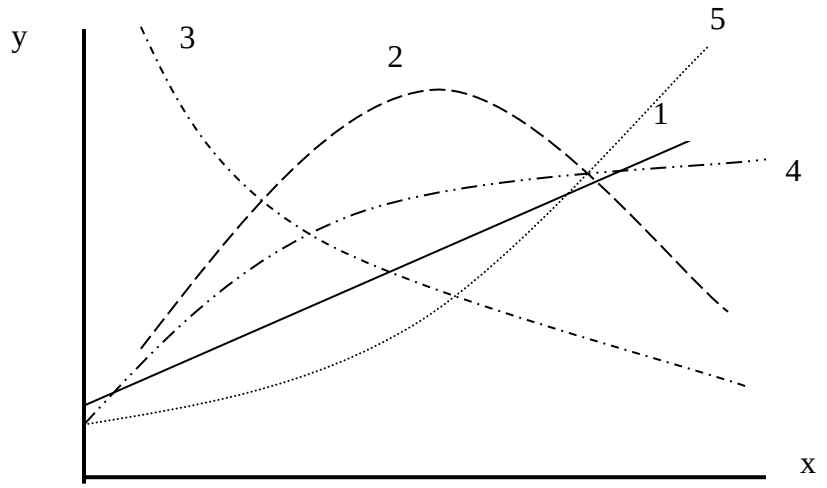


Рис.1. Лінії, які характеризують різні типи залежностей:

——— - лінійна залежність;

– – – – – - параболічна залежність;

- - - - - - - - - - - - гіперболічна залежність;

– - - – - - – - степенева залежність;

..... - показникова залежність.

Кожному виду залежності відповідає своє математичне рівняння. Так, лінійна залежність (лінія 1) описується рівнянням:

$$y = a + b \cdot x, \tag{11}$$

де a, b, c - постійні коефіцієнти;

x - незалежна змінна;

- параболічна залежність (лінія 2):

$$y = a + b \cdot x + c \cdot x^2, \tag{12}$$

- гіперболічна залежність (лінія 3):

$$y = a + \frac{b}{x}, \tag{13}$$

- степенева залежність (лінія 4):

$$y = a \cdot x^b, \tag{14}$$

- показникова залежність (лінія 5)

$$y = a \cdot b^x. \tag{15}$$

Побудова математичної моделі, яка б найбільш точно описувала залежність між досліджуваними ознаками, надалі, полягає в знаходженні параметрів a, b, c вибраного типу рівняння.

Повернутись до плану.

Подивитись навчальні відео

https://www.youtube.com/watch?v=XbufBZnq3oo&t=347s

https://www.youtube.com/watch?v=nlvq-L3s5Lc