УДК: 371.26

Павло ВОЛОВИК
м. Київ

Педагогічна технологія застосування регресійного і кореляційного аналізів в педагогічних дослідженнях

У статті розглядається суть та можливості регресійного і кореляційного аналізів при дослідженні педагогічних явищ та процесів; технологія виявлення взаємозв'язків між педагогічними явищами та процесами, знаходження їх кількісної оцінки, встановлення закономірностей взаємозв'язаних педагогічних явищ і показників, що їх характеризують; розкривається методика побудови емпіричних закономірностей (емпіричних формул) на основі дослідних даних, зокрема із застосуванням методів: натянутої нитки, вибраних точок, найменших квадратів (метода Гаусса), інтерполяційної формули Лагранжа та ін. У статті також висвітлюється методика побудови кореляційних рівнянь (рівнянь регресії), розкривається суть та методика визначення емпіричних мір тісноти зв'язку (коефіцієнта взаємної спряженості, коефіцієнта кореляції та ін.).

Матеріалістична діалектика, як відомо, вчить, що всі предмети і явища існують не самі по собі, не ізольовано, а в нерозривному взаємозв'язку, єдності з іншими предметами і явищами. Кожний з них діє на інші предмети і явища й зазнає впливу з їх боку.

Загальний зв'язок і взаємна обумовленість предметів і явищ становлять істотну особливість матеріального світу. Тому одним з найважливіших завдань всякого дослідження, зокрема і педагогічного, є встановлення зв'язку між величинами або факторами, зміна яких визначає сутність процесу, що вивчається. Зв'язки в навколишньому світі дуже різноманітні й складні. Щоб пізнати яке-небудь явище, треба вивчити не тільки його зв'язки з навколишніми явищами - факторами, але також взаємозв'язки всіх його сторін, тобто треба встановити закономірності змін взаємопов'язаних явищ і показників, що їх характеризують.

У математичній статистиці взаємозв'язок явищ вивчається методом кореляції. Потреба в застосуванні цього методу зумовлюється тим, що не завжди можна врахувати вплив сторонніх факторів, або через те, що ці фактори невідомі, або через те, що їх не можна ізолювати. Метод кореляції дає змогу при складних взаємодіях сторонніх впливів з'ясувати, яка була б залежність між результатом і фактором, якби сторонні фактори не змінювались і своєю зміною не спотворювали основну залежність. Для виявлення закономірності зв'язку треба мати достатньо велике число спостережень.

За допомогою кореляції розв'язують такі дві задачі:

1) на основі спостережень над великим числом фактів визначають як змінюється в середньому результуюча ознака із зміною певного факто-

ра, припускаючи при цьому, що інші фактори не змінюються, хоч насправді спотворюючий їх вплив має місце;

2) визначають ступінь впливу спотворюючих факторів.

Перед тим як розглядати кореляційний зв'язок між двома змінними x і y розглянемо, як знаходити емпіричні залежності між дослідними залежними величинами, коли кожному значенню незалежної ознаки x відповідає лише одне значення залежної ознаки y.

1. Емпіричні формули

У процесі вивчення багатьох питань природознавства, економіки, педагогіки часто доводиться на основі великої кількості дослідних даних встановлювати кількісні залежності між різними зв'язаними одна з одною величинами (ознаками). **Такі залежності називаються емпіричними.**

Побудова емпіричних залежностей або емпіричних формул полягає в такому. Нехай в результаті спостережень ми дістали два ряди значень для деякої незалежної ознаки x і залежної від неї ознаки y, а саме: встановили, що коли незалежна ознака x набуває значень $x_1; x_2; ...; xn$, то залежна ознака y набуває відповідно значень $y_1; y_2; y_3; ...; yn$. Треба знайти функцію $y = f(x)$, яка добре відображала б нашу таблицю дослідних даних:

x	⟵	$x_1;$	$x_2;$	$...; xn$
		↓	↓	↓
y	⟵	$y_1;$	$y_2;$	$...; yn$

Інакше кажучи, слід знайти таку функцію $y = f(x)$, яка при значеннях незалежної ознаки $x_1; x_2; ...; xn$ давала б значення $f(x_1); f(x_2); ...; f(xn)$, досить близькі до табличних значень залежної ознаки $y_1; y_2; ...; yn$.

Щоб легше було встановити аналітичну залежність між величинами на площині, будуємо точки з координатами $(x_1; y_1); (x_2; y_2); ...; (xn; yn)$. Сполучивши послідовно ці точки, дістанемо ламану лінію, що є графічним зображенням табличних даних.

Після цього дивимося, яка з відомих нам кривих більше підходить до побудованої ламаної, тоді для вираження залежності між x і y беремо рівняння цієї кривої $f(x)$. Ця операція називається згладжуванням ламаної за допомогою кривої.

При цьому намагаємося підібрати найпростіші формули: прямої лінії, параболи другого або третього порядку, гіперболи і показникової функції.

Розглянемо деякі методи побудови емпіричних формул.

2. Метод натягнутої нитки

Нехай в результаті спостережень ми дістали два ряди значень для незалежної і залежної ознак:

x	⟵	$x_1; x_2; ...; xe; ...; xk; ...; xn;$
y	⟵	$y_1; y_2; ...; ye; ...; yk; ...; yn.$

Побудуємо на площині точки *(x1; y1); (x2; y2); ...; (xe; ye); ...; (xk; yk); ...;* (xn; yn). Якщо вони будуть розміщені у вигляді вузького снопа (рис. 1), то вибираємо дві точки A *(xe; ye)* і B *(xk; yk)* і проводимо через них пряму лінію (натягуємо нитку). Точки *A* і *B* вибираємо так, щоб по обидві сторони прямої *AB* розмістилась приблизно однакова кількість точок і щоб вони, по можливості, були ближче до прямої AB.

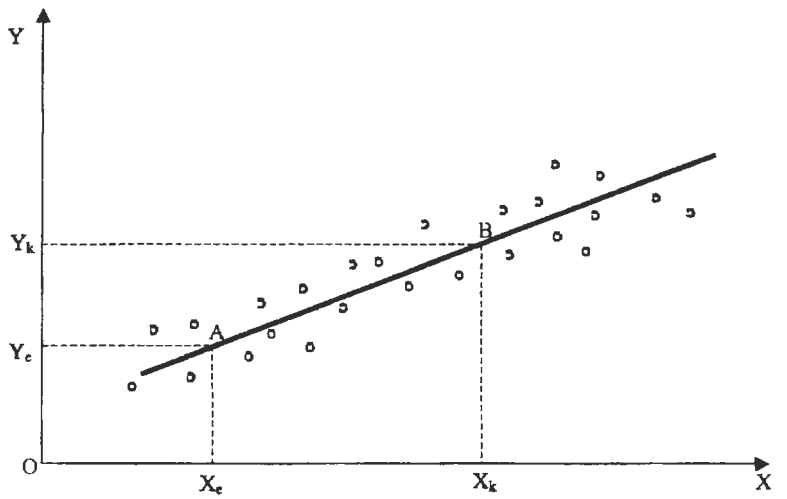

рис. 1.

Після цього складаємо рівняння прямої, яка проходить через дві точки *A (xe; ye); B (xk; yk)*:

$$\frac{x-x_e}{x_k-x_e}=\frac{y-y_e}{y_k-y_e} \qquad (1)$$

Розв'язавши його відносно y , дістанемо:

$$(x - x_e)\,(y_k - y_e) = (x_k - x_e)\,(y - y_e);$$

$$x\,(y_k - y_e) - x_e\,(y_k - y_e) = y\,(x_k - x_e) - y_e\,(x_k - x_e);$$

$$y=-x_e\frac{y_k-y_e}{x_k-x_e}+y_e\frac{x_k-x_e}{x_k-x_e}+\frac{y_k-y_e}{x_k-x_e}x.$$

Ввівши позначення - $\frac{x_e(y_k-y_e)}{x_k-x_e}+y_e=a$ і $\frac{y_k-y_e}{x_k-x_e}=b$, дістанемо рівняння $y=a+bx$,

яке зображає приблизну лінійну залежність між *x* і *y*.

Приклад. Зростання кількості міського населення на Україні (в процентах до 1913 р.) характеризується такими даними[1]:

[1]Україна за п'ятдесят років (1917 - 1967). Статистичний довідник. Держполітвидав України. - К., - 1967. - С. 28.

Таблиця 1

Роки	1913	1939	1959	1965	1966
Міське населення в процентах	100	200	282	349	357

Побудувати емпіричну формулу, застосувавши метод натянутої нитки, яка характеризує зміну кількості міського населення в Україні в період 1913-1966 рр.

Розв'язання. Насамперед будуємо систему координат і на координатну площину наносимо точки з такими координатами А (1913, 100); В (1939, 200); С (1959, 282); *D* (1965, 349) і *E* (1966, 357) (рис. 2).

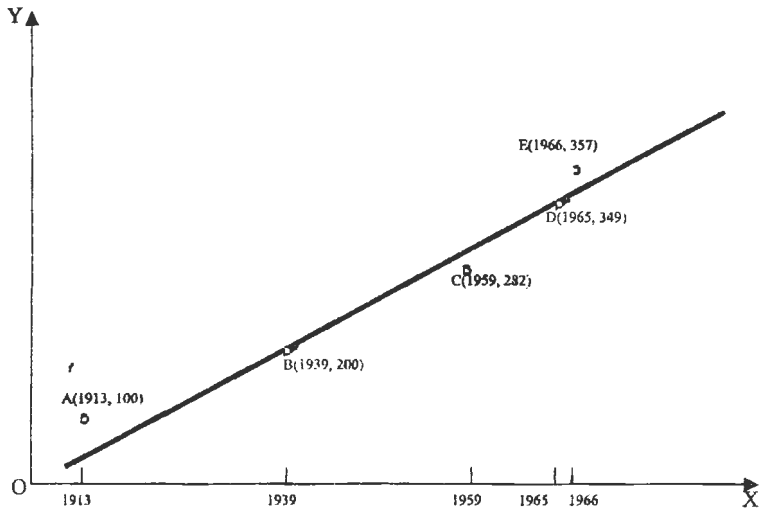

рис. 2.

Вибираємо дві точки. В нашому випадку доцільно взяти точки *B* і *D*, оскільки від прямої, проведеної через ці точки, решта точок будуть по обидві сторони, причому на незначній відстані.

За формулою (1) визначаємо *y*:

$$y = \frac{-x_B(y_D - y_B)}{x_D - x_B} + y_B + \frac{y_D - y_B}{x_D - x_B} * x = -\frac{1939(349-320)}{1965-1989} + 200 + \frac{349-200}{1965-1939} * x \approx -10912 + 5{,}73x$$

Отже, залежність, що характеризує зміну міського населення в Україні в період 1913-1966 рр., визначається приблизно такою емпіричною формулою:

$$y = -\ 10912 + 5{,}73*x\ ,$$

де x - календарний рік.

3. Метод вибраних точок

Метод натягнутої нитки, який ми розглянули вище, є частковим випадком загальнішого методу, що називається методом вибраних точок. Суть його полягає в такому. Дослідні дані $x_1; x_2; ...; x_n$ і значення $y_1; y_2; ...; y_n$, що їм відповідають, наносимо на координатну площину у вигляді точок $(x_1; y_1); (x_2; y_2); ...; (x_n; y_n)$. Вибираємо таку криву, щоб точки розміщалися близько біля неї. Нехай це буде парабола другого порядку. Тоді залежність між x і y шукаємо у вигляді квадратної функції $y = a + bx + cx^2$. Щоб знайти коефіцієнти a, b, c вибираємо три точки, через які проходить парабола. Координати цих точок підставляємо по черзі в рівняння $y = a + bx + cx^2$. При цьому дістаємо систему трьох рівнянь, розв'язавши яку матимемо крефіцієнти a, b, c.

Якщо емпіричні точки розміщаються ближче до кривої третього порядку $y = a + bx + cx^2 + dx^3$, або до гіперболи $y = a + \frac{b}{x}$, або до якої-небудь іншої кривої, то вибираємо стільки опорних точок, скільки коефіцієнтів у рівнянні. Підставивши значення цих точок у рівняння, матимемо систему рівнянь, розв'язавши яку, визначимо коефіцієнти.

Вибираючи криву, треба добиватись, щоб більшість точок знаходилась поблизу кривої $f(x)$.

Приклад. Зростання кількості студентів вищих навчальних закладів по роках в Українській РСР характеризується такими даними[2]:

Таблиця 2

На початок навчального року (x)	1940/41	1950/51	1960/61	1966/67
Кількість студентів вузів у тис. чол. (y)	196,8	201,5	417,7	739,1

Побудувати емпіричну формулу вигляду $y = a + bx + cx^2$, яка характеризує і кількість студентів вузів на початок навчального року (y) в період 1940-41 - 1966-67 н.р.

Розв'язання. З усіх можливих парабол $y = a + bx + cx^2$ вибираємо ту, яка проходить через точки, що відповідають 1950/51 навчальному року $x_3 = 10$ і $y_1 = 201,5$; 1960/61 н.р. $x_2 = 20$ *і* $y_2 = 417,7$; 1966/67 н.р. - $x_3 = 27$ *і* $y_3 = 739,1$ (рис. 3).

---

[2] Україна за п'ятдесят років (1917-1967). Статистичний довідник. Держполітвидав України. - К., - 1967. - С. 216.

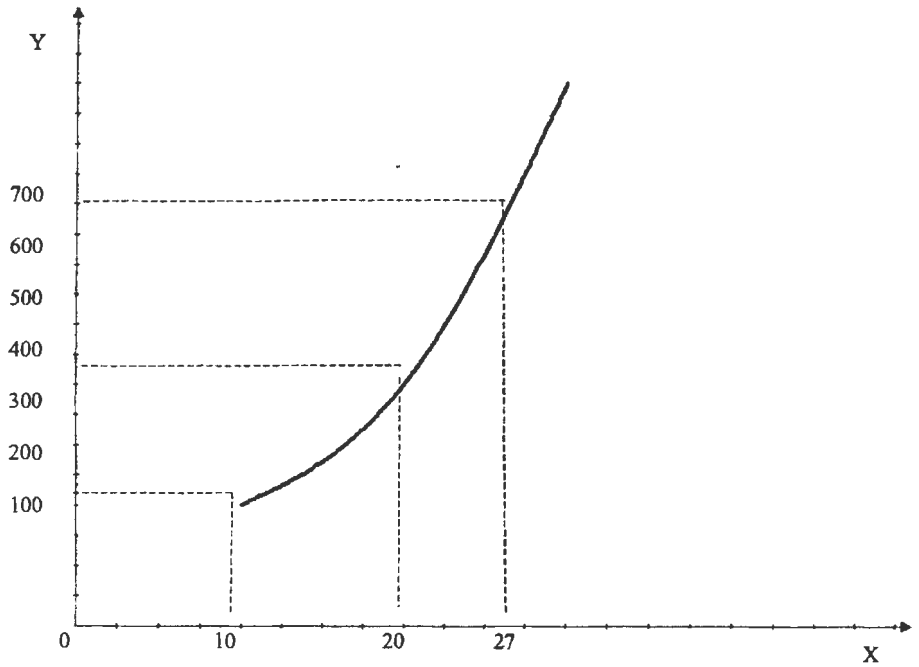

рис. 3.

Підставивши значення координат цих точок у рівняння $y = a + bx + cx^2$, дістанемо систему трьох рівнянь з трьома невідомими:

$$201{,}5 = a + 10b + 100c;$$
$$417{,}7 = a + 20b + 400c;$$
$$739{,}1 = a + 27b + 729c;$$

Розв'язавши цю систему, дістанемо: $a = 271{,}3$; $b = -21{,}28$; $c = 1{,}43$. Підставимо значення a, b і c в рівняння $y = a + bx + cx^2$, одержимо емпіричну формулу $y = 271{,}3 - 21{,}28x + 1{,}43x^2$, яка характеризує зростання кількість студентів на початок навчального року в період 1940/41 - 1966/67 н.р.

4. Інтерполяційна формула Лагранжа

Коли емпіричну формулу треба знайти у вигляді многочлена вище як другого степеня, то часто користуються так званою формулою Лагранжа.

Нехай дані нашого досліду зведені в таблицю

x	x_1	x_2	...	x_n
y	y_1	y_2	...	y_n

Треба знайти многочлен n - I степеня $y=\phi(x)$, який при заданих значеннях незалежної ознаки $x_1; x_2; ...; x_n$ набував би відповідно здобутих в результаті досліду значень залежної ознаки $y_1; y_2; ...; y_n$, тобто щоб виконувалась така система рівнянь:

$$\left.\begin{aligned}
&\phi(x_1) = y_1,\\
&\phi(x_2) = y_2,\\
&\cdots\cdots\cdots\cdots\\
&\cdots\cdots\cdots\cdots\\
&\phi(x_n) = y_n.
\end{aligned}\right\}$$

Цей поліном шукаємо так. Нехай шуканим многочленом буде

$$y = \phi(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_{n-1} x^{n-1}$$

,

причому a_0 , a_1 ., a_2 , ..., a_{n-1} - деякі невідомі коефіцієнти. Підставляємо в праву частину рівняння замість х по черзі значення x_1; x_2; ...; x_n. За прийнятим вище припущенням многочлен повинен набути відповідних значень y_1; y_2; ...; уп. При цьому дістанемо систему п рівнянь:

$$a_0 + a_1 x_1 + a_2 x_1^2 + \ldots + a_{n-1} x_n^{n-1} = y_1,$$
$$a_0 + a_1 x_2 + a_2 x_2^2 + \ldots + a_{n-1} x_2^{n-1} = y_2,$$
$$\cdots\cdots\cdots\cdots$$
$$\cdots\cdots\cdots\cdots$$
$$a_0 + a_1 x_n + a_2 x_n^2 + \ldots + a_{n-1} x_n^{n-1} = y_n.$$

Розв'язавши цю систему, знайдемо коефіцієнти a_0 , a_1 ,a_2 , ... , a_{n-1} . Підставимо їх значення в многочлен:

$$\phi(x) = \frac{(x-x_2)(x-x_3)\ldots(x-x_n)}{(x_1-x_2)(x_1-x_3)\ldots(x_1-x_n)} y_1 +$$
$$+ \frac{(x-x_1)(x-x_3)\ldots(x-x_n)}{(x_2-x_1)(x_2-x_3)\ldots(x_2-x_n)} y_2 + \ldots +$$
$$+ \frac{(x-x_1)(x-x_2)\ldots(x-x_{n-1})}{(x_n-x_1)(x_n-x_2)\ldots(x_n-x_{n-1})} y_n.$$

Якщо в цю формулу замість х підставити x_1 , то дістанемо $\phi(x_1) = y_1$, оскільки решта членів дорівнюватиме нулю, бо в їх чисельники входять співмножники $(x - x_1)$, які, коли підставити $x = x_1$, дорівнюють нулю. Відповідно, якщо замість х підставити x_2 , x_3 , ... , x_n , то $\phi(x_2) = y_2$; $\phi(x_3) = y_3$; ... $\phi(x_n) = y_n$.

Формула (2) називається інтерполяційною формулою Лагранжа. Інтерполяційний характер цієї формули полягає в тому, що її часто використовують для обчислення проміжних значень залежної ознаки у , тобто значень у , які відповідають проміжним значенням х , що не задані таблицею і знаходяться, наприклад, між x_1 і x_2 ; x_2 і x_3 і т.д.

Приклад. У табл. 3 наведено дані про кількість учнів V - VIII класів у загальноосвітніх школах Української РСР. Встановити закономірність зміни кількості учнів V - VIII класів.

Таблиця 3.

На початок навчального року	1940/41	1950/51	1960/61	1966/67
Кількість учнів V – VIII класів у тис.	2525,5	3053,0	2949,6	3350,7

Розв'язування. Взявши 1940/41 навчальний рік за початок відліку і позначивши число років, які пройшли через x , складаємо допоміжну таблицю (табл. 4).

Таблиця 4.

x	1	10	20	27
y	2525,5	3053,0	2949,6	3350,7

За формулою Лагранжа знаходимо многочлен 3-го степеня, який при x = 1, 10, 20, 27 давав би відповідно значення y = 2525,5; 3053,0; 2949,6; 3350,7.

Отже, між кількістю учнів у V - VII класах загальноосвітніх шкіл Української РСР (в тисячах) і роками, починаючи з 1940/41 навчального року, існує така емпірична залежність: $y = 0{,}30x^3 - 13{,}20x^2 + 175{,}16x + 2327{,}40$.

За здобутою емпіричною залежністю можна обчислювати проміжні значення кількості учнів у цих класах. Наприклад, щоб визначити, скільки було учнів у V - VIII класах в 1955/56 навчальному році, підставляємо в емпіричне рівняння значення x = 15, y = 0,30 * 153 - 13,20 * 152 + 175,16 * 15 + 2327,40 1012,5 - 2970,0 + 2627,4 + 2327,40 = 3097,3 (тис.).

5. Метод найменших квадратів (метод Гаусса)

Нехай в процесі певного дослідження ми дістали такі дані:

x	x_1	x_2	x_3	...	x_n
y	y_1	y_2	y_3	...	y_n

Виходячи із змісту розглядуваних явищ, припускаємо, що між цими величинами існує певна функціональна залежність $y(x)$. Метод найменших квадратів полягає в тому, що треба знайти такі параметри функціональної залежності $y(x)$, щоб сума квадратів відхилень фактичних (дослідних) даних від вирівняних була найменшою (рис. 4).

$S = (y_1 - y_1)^2 + (y^2 - y^2)^2 + ... + (y_n - y_n)^2,$ (3)

де y_1 - фактичні (дослідні) значення;

y_1 - вирівняні значення.

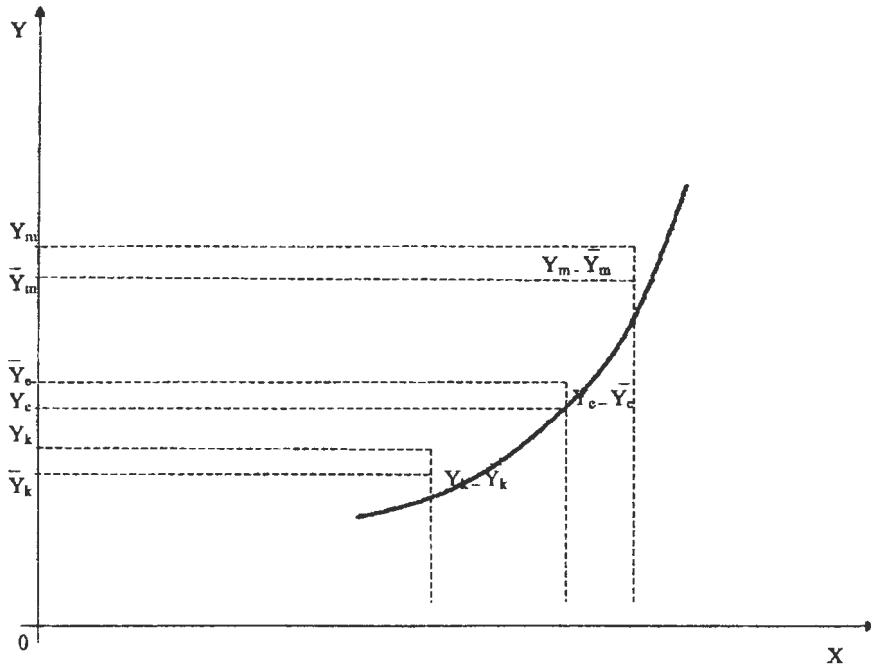

рис. 4.

Застосуємо цей метод для визначення параметрів деяких конкретних функціональних залежностей.

а) Нехай між даними табл. 4 існує прямопропорціональна залежність, тобто теоретична крива, за допомогою якої будемо вирівнювати емпіричну залежність між цими величинами, має такий вигляд:

$y = kx.$ (4)

Тоді (3) запишеться ŷ вигляді

$$S = \sum_{i=1}^{n} (y_i - kx_i)^2.$$

Як видно, ця сума залежить від k. Вона буде мінімальна тоді, коли похідна по змінній k дорівнює нулю, тобто

$$\frac{dS}{dk} = 2\sum_{i=1}^{n}(y_i - kx_i)(-x_i) = 0.$$

скоротимо це рівняння на -2:

$$\sum_{i=1}^{n}(y_i - kx_i)x_i = 0; \sum_{i=1}^{n} x_i y_i - k\sum_{i=1}^{n} x_i^2 = 0;$$

$$k\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i,$$

звідки

$$k = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

Підставивши значення k в рівняння (4), дістанемо:

$$y = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} * x. \qquad (4a)$$

б) Нехай функціональна залежність має такий вигляд:

$$y = a + bx.$$

Підставивши в рівняння (3) замість y_i відповідно $a + bx_i$, дістанемо

$$S = \sum_{i=1}^{n} (y_i - a - bx_i)^2.$$

У цій функції невідомі коефіцієнти a і b, при яких функція матиме мінімальне значення. Щоб знайти ці значення, візьмемо частинні похідні по a та b і прирівняємо їх до нуля. Розв'язок здобутої системи рівнянь дає ті значення, при яких дана сума мінімальна.

$$\left.\begin{aligned} \frac{ds}{da} &= 2\sum (y_i - a - bx_i)(-1) = 0, \\ \frac{ds}{db} &= 2\sum (y_i - a - bx_i)(-x_i) = 0. \end{aligned}\right\}$$

Скоротимо обидва рівняння на -2 і зробимо такі перетворення:

$$\left.\begin{aligned} \sum_{i=1}^{n} (a + bx_i) &= \sum_{i=1}^{n} y_i, \\ \sum_{i=1}^{n} (ax_i + bx_i^2) &= \sum_{i=1}^{n} x_i y_i, \end{aligned}\right\};$$

$$\left.\begin{aligned} \sum_{i=1}^{n} a + b\sum_{i=1}^{n} x_i &= \sum_{i=1}^{n} y_i, \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} (x_i y_i). \end{aligned}\right\}$$

Врахувавши, що $\sum_{i=1}^{n} a = na$, дістанемо:

Опустивши індекси, перепишемо систему (5) так:

$$\left.\begin{aligned} na + b\sum_{i=1}^{n} x_i &= \sum_{i=1}^{n} y_i, \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} x_i y_i. \end{aligned}\right\} \qquad (5)$$

Здобута система рівнянь називається нормальною системою Гаусса. Розв'язавши її, знайдемо значення a і b:

$$a=\frac{\begin{vmatrix}\sum y & \sum x\\ \sum xy & \sum x^2\end{vmatrix}}{\begin{vmatrix}n & \sum x\\ \sum x & \sum x^2\end{vmatrix}}=\frac{\sum y*\sum x^2-\sum xy*\sum x}{n\sum x^2-(\sum x)^2}; \qquad (6)$$

$$b=\frac{\begin{vmatrix}n & \sum y\\ \sum x & \sum xy\end{vmatrix}}{\begin{vmatrix}n & \sum x\\ \sum x & \sum x^2\end{vmatrix}}=\frac{n\sum xy-\sum x\sum y}{n\sum x^2-(\sum x)^2}. \qquad (7)$$

в) Аналогічно складається система нормальних рівнянь, коли зв'язок між ознаками близький до оберненого і досить добре виражається залежністю

$$y=a+\frac{b}{x}.$$

Система нормальних рівнянь для цього випадку буде така:

$$\left.\begin{aligned} na+b\sum\frac{1}{x}&=\sum y;\\ a\sum\frac{1}{x}+b\sum\frac{1}{x^2}&=\sum\frac{y}{x}.\end{aligned}\right\} \qquad (8)$$

6. Екстраполяція

В практиці часто доводиться складати прогнози розвитку того чи іншого явища за дослідними даними за умови, що тенденція розвитку явища не змінюється. Ця операція називається екстраполяцією, тобто знаходженням наступних рівнів ознаки, коли попередні відомі.

Наприклад, коли ми, вирівнюючи рівні якого-небудь ряду за параболою 2-го порядку, дістали рівняння зв'язку $y=a+bx+cx^2$ і маємо підставу припускати, що і в наступні роки ця тенденція не зміниться, то, підставивши в рівняння зв'язку значення x, дістанемо прогнози значення ознак.

Приклад. За емпіричною формулою, яка характеризує кількість студентів вузів на початок навчального року (див. приклад п.3) знайти, скільки буде студентів в Україні у 1975 р., припустивши, що тенденція зростання лишиться такою самою, як і раніше.

Розв'язання. Підставивши в рівняння зв'язку $y = 271,3 - 21,28x + 1,43x^2$ значення $x = 35$, дістанемо:

$$y=271,3-21,28*35+1,43*35^2=1278,25 \quad \text{(тис.)}.$$

Слід зазначити, що завдання, пов'язані з відшукуванням емпіричної формули, яка добре відображає зв'язок між ознаками x і y, не є визначеними. Можна скласти нескінченну множину залежностей $y = f(x)$, які відповідали б даній меті. Це означає, що функцію $y = f(x)$ ми вибираємо самі, але при цьому намагаємось, щоб з усіх можливих залежностей, які досить добре відображають емпіричні дані, вибрано було найпростішу.

Продовження статті в наступному номері журналу.

Стаття надійшла до редакції 07.02.07

Павло ВОЛОВИК

Педагогическая технология применения регрессионного и корреляционного анализов в педагогических исследованиях

Резюме

В статье рассматриваются сущность и возможности регрессионного и корреляционного анализов при исследовании педагогических явлений и процессов; технология выявления взаимосвязей между педагогическими явлениями и процессами, нахождения их количественной оценки, установления закономерностей взаимосвязанных педагогических явлений и показателей, характеризующих их; раскрывается методика построения эмпирических закономерностей (эмпирических формул); на основе опытных данных, в частности с применением методов: натянутой нити, избранных точек, наименьших квадратов (метода Гаусса), интерполяционной формулы Лагранжа и др. В статье также излагается методика построения корреляционных уравнений (уравнений регрессии), раскрывается сущность и методика определения эмпирических мер тесноты связи (коэффициента ассоциации, коэффициента взаимной сопряженности, коэффициента корреляции и др.).

Pavlo VOLOVYK

The Pedagogical Technology of Regression and Correlation Analysis Usage in Pedagogical Researches

Summary

The crux and resources of regression and correlation analysis at research of the pedagogical phenomena and process are considered in this article. Technology of revealing of correlations between pedagogical process and its phenomena is examined here. Determination of its quantitative estimation, fixation of correlated patterns between pedagogical phenomena and its characterized index are determined in this article. Methods of empirical patterns formation (empiric formulas) are considered here. It is carried out investigation on the basis of studied data. Usage of such methods as: strained thread, selected points, the least squares (method of Gauss), La Grange interpolation formula etc. can be regarded here. The technique of correlation equations' (the equation of regression) construction is stated in this article. One can also find the revelation of the empirical measures of connection narrowness essence and its technique (association factor, factor of reciprocal conjugacy and factor of correlation).